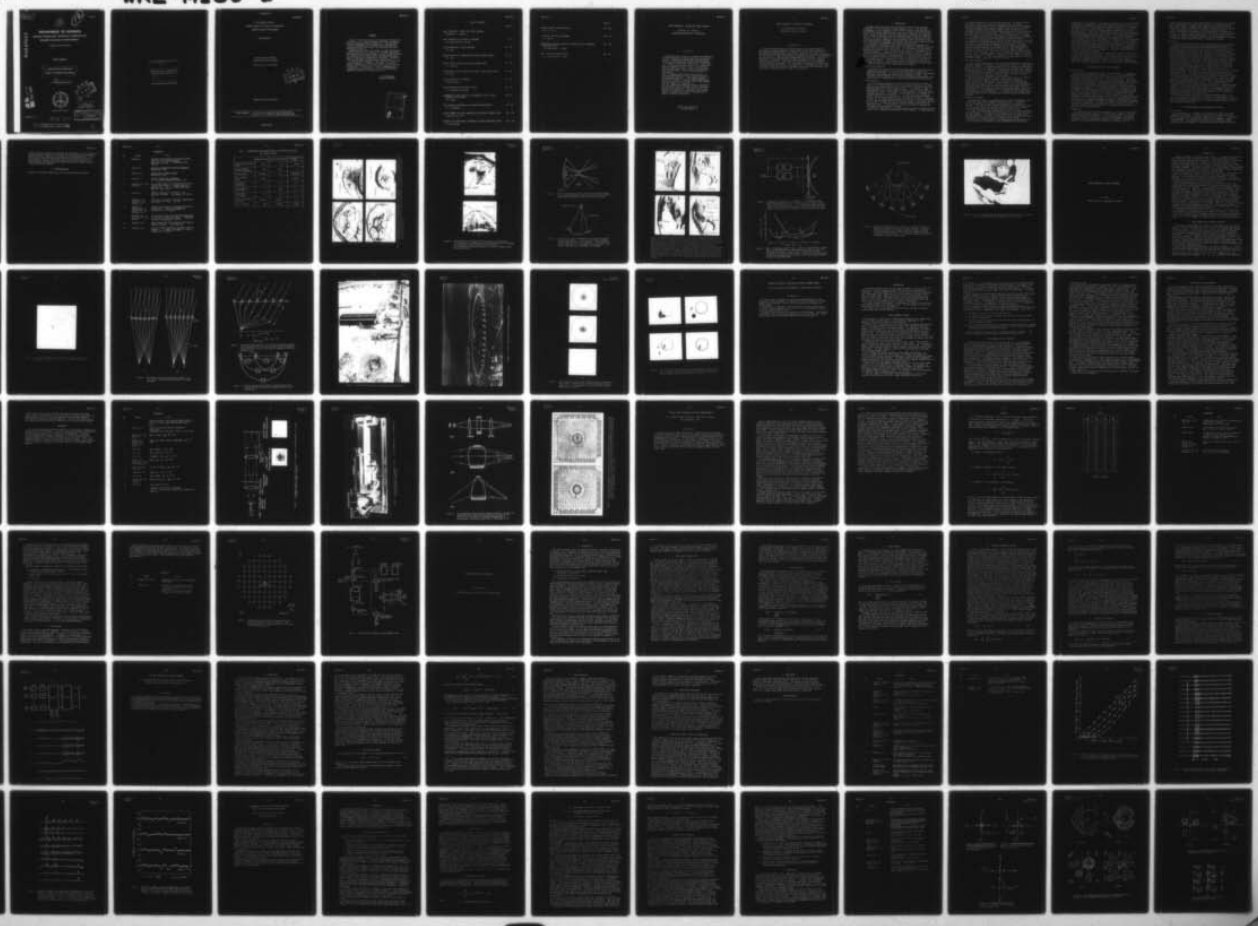
AD-A047 017 WEAPONS RESEARCH ESTABLISHMENT SALISBURY (AUSTRALIA) F/G 12/1
PROCEEDINGS OF SYMPOSIUM SIGNAL PROCESSING FOR ARRAYS,(U)
MAY 77 H A D'ASSUMPCAO

UNCLASSIFIED WRE-MISC-1 NL

1 OF 3
ADA
047017



(14) WRE-MISC-1

AR-000-564

(12) B.S.

AUSTRALIA

DEPARTMENT OF DEFENCE

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION

WEAPONS RESEARCH ESTABLISHMENT

SALISBURY, SOUTH AUSTRALIA

MISCELLANEOUS 1

(6) PROCEEDINGS OF SYMPOSIUM

SIGNAL PROCESSING FOR ARRAYS

(10) EDITED BY H.A. D'ASSUMPCAO

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION

(12) 242p.

Approved for Public Release

COPY No. 178

371 700

ORIGINAL CONTAINS COLOR PLATES: ALL DDC REPRODUCTIONS WILL BE IN BLACK AND WHITE.

AB

AR-000-564

DEPARTMENT OF DEFENCE

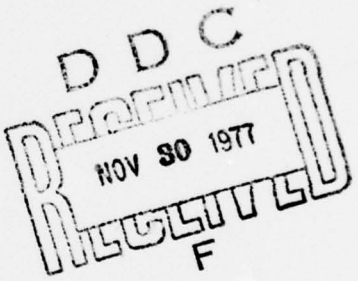
DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION

WEAPONS RESEARCH ESTABLISHMENT

MISCELLANEOUS 1

PROCEEDINGS OF SYMPOSIUM
SIGNAL PROCESSING FOR ARRAYS

Edited by H.A. d'Assumpcao



Approved for Public Release.

POSTAL ADDRESS: The Director, Weapons Research Establishment,
Box 2151, G.P.O., Adelaide, South Australia, 5001.

A

FOREWORD

Arrays of receivers are used in many fields of science and engineering - e.g., radio astronomy, seismology, underwater acoustics, radar and medical ultrasonics. Although there are great differences in the hardware for processing the signals received by such arrays, the underlying principles are generally similar.

Recognising that there are a number of people working with various types of arrays in Australia, a two-day Symposium at the Weapons Research Establishment, Salisbury, South Australia, was sponsored by the Defence Science and Technology Organization of the Australian Department of Defence. There were about 100 attendees, including several from overseas.

In editing the Proceedings of this Symposium, I have been greatly assisted by a number of colleagues. I wish to express my appreciation to them and to the authors for their co-operation. In particular, I wish to acknowledge the invaluable help of Professor R.G. Keats of the Mathematics Department, University of Newcastle, N.S.W.

H.A. d'Assumpcao
Symposium Convener

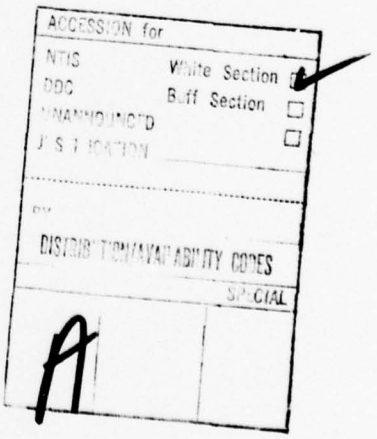


TABLE OF CONTENTS

ARRAY PROCESSING: PUTTING THE PIECES TOGETHER

Professor P.L. Stocklin
Loughborough University of Technology

S U M M A R Y

The purpose of this paper is to outline and demonstrate the common principles of modern approaches to array processing. Technological developments associated with several of these approaches are discussed and compared.

The paper is divided into three parts. In the first, Part I: Conceptual, decision theory in its sampled and amplitude-quantized version is used to develop basic array processing principles, primarily the structures used for processing. Criteria and performance are illustrated.

Part II: Examples demonstrates these principles in classical beam formation, spatial adaptive processing and superdirective arrays. The common principle behind these examples is formally demonstrated and the characteristics of the noise and signal fields leading to these examples are contrasted.

Part III: Supporting Technology summarizes constraints and approaches in the technology of array processing. Low-bit, delta modulation, transversal filter, Gram-Schmidt and convolutional array methods are included, together with discussion of matrix inversion as it applies to adaptive processing.

(Paper not available on publication)

ARRAY TECHNIQUES IN DIAGNOSTIC ULTRASOUND

D.E. Robinson and G. Kossoff,

Ultrasonics Institute, Sydney

SUMMARY

The ultrasonic pulse-echo visualisation technique has found widespread application in medical diagnostics, particularly in soft tissued areas such as the abdomen and pregnant uterus, eye, breast, thyroid and the infant brain. The similarities and differences between the imaging for the ultrasonic technique and that for radar, sonar and exploration seismology are outlined; in particular, the differences in signal processing requirements are discussed. Three new departures in ultrasonic visualisation under development at the Ultrasonics Institute are described. These are a multi-transducer array for improved clinical operation, an annular array for diagnostic focusing and a linear array for beam steering and focusing. The application of two-dimensional processing techniques, using either the multi-transducer or the wide aperture arrays, and their role in future ultrasonic systems, are discussed.

1. INTRODUCTION

Ultrasonic pulse echo visualisation techniques have received widespread acceptance in clinical practice in many diagnostic areas. They are used predominantly in the examination of soft tissues and have found application in the abdominal organs, pregnant uterus, heart, eye, thyroid, breast and infant brain. As the resolution, sensitivity and display techniques have improved, the technique has been used in an ever widening range of areas and conditions (ref.1). The aim of this paper is to describe the basic properties of this visualisation method, compare it with the somewhat similar techniques of radar, sonar and exploration seismology, and describe some of the processing being used and in the course of development at the Ultrasonics Institute.

In ultrasonic visualisation, short pulses of energy are propagated in a narrow beam into the examined volume, echoes are returned by specular reflecting surfaces and scatterers within the tissue. The time delays of these echoes depend on the distance to the reflector and the velocity of propagation in the intervening medium. A cross sectional display similar to a radar PPI can be obtained by moving the transducer line of sight around the part to be examined within a single plane with a scanning motion. A line on the CRT screen is moved in a similar fashion and echoes displayed as intensification of the line may be stored either in a storage tube, a scan convertor tube or directly on film with time exposure. The display resembles an anatomical cross section with bright outlines for reflecting surfaces between organs and textured echoes from within structures with scattering properties (figure 1).

1.1 Pulse echo techniques

Superficially, the ultrasonic technique would appear to have much more in common with radar and active sonar, and somewhat less in common with exploration seismology. In fact the basic differences between the techniques are quite large and probably exploration seismology is closest. The differences can be examined under the general headings of beam pattern, echo signal properties and properties of the transmission medium and reflectors.

1.2 Beam patterns

The beam pattern parameters of typical pulse-echo systems are shown in Table 1. The range to aperture ratio is a measure of the ability to focus the system and is equivalent to the f number or the reciprocal of twice the numerical aperture in a focused optical system. As in optics, a small range/aperture ratio or large numerical aperture implies a sharp focus and a small depth of field. The aperture to wavelength ratio is a measure of the azimuthal resolution of the system. It also influences the diffraction limited resolution of a focused system. A large value reduces the beam width in the far field and improves the resolution at the focus. The ratio of range to near field length (given by $A^2/4\lambda$, where A is the linear aperture and λ is the wavelength) indicates whether the system is operating in the near or far field. A value less than one indicates operation in the near or diffraction field. In this range, an increase in aperture increases the beam width unless the system is focused.

Sonar(ref.2) and radar(ref.3) systems operate in the far field and the beam pattern can be adequately represented by a polar diagram. A larger aperture

and a higher frequency give rise to a narrower beam. The target distance to aperture ratio is very large and no focusing action is possible. Exploration seismology(ref.4) uses an expanding wavefront with beam-forming from received signals from an array, usually a linear array of 24 elements. The depth to aperture ratio allows reasonable focus, the aperture to wavelength ratio implies that a narrow beam and sharp focus are possible and the system operates in the near field. The transmitter is a point source and focusing is carried out by processing of the received data. Although it is not thought of in this way, the procedure of normal move-out correction and stacking is equivalent to one dimensional focusing on reception.

In ultrasonic systems(ref.5), the aperture to wavelength ratio is typically some hundreds, allowing good beaming and the distance to aperture ratio is in the range 2 to 20 permitting a sharp focus to be obtained if required. In a typical ultrasonic system for examination of the abdomen and pregnant uterus using water coupling, the frequency used is 2 MHz (wavelength of 0.75 mm) the transducer aperture A is 70 mm and the range of depths to be examined is 200 to 400 mm. The length of the near field is given by $A^2/4\lambda$ and is 1.3 m in the sample system. Thus the operating range is well within the near field. The transducer is focused with a focal length F of 280 mm, giving an f number of 4. The 20 dB beam width or resolution at the focus on a reflection basis is given by $1.74\ \lambda F/A$ or 5.2 mm.

1.3 Echo signal properties

Radar, sonar and exploration seismic systems are generally all peak-power and signal/random noise limited. To achieve greater range it is possible to extend the length of the transmitted pulse, thus increasing the signal power. However this reduces the range resolution which may be regained by a variety of deconvolution filter techniques. Random incoherent noise can be reduced by matched filtering and pulse-to-pulse averaging. The ultrasonic system is not peak power or random noise limited and this complexity of signal processing has not been required up to this stage.

The axial or depth resolution obtained with a short pulse is in practice several times better than the azimuthal or beam width resolution. Processing currently employed uses time varying gain to compensate for attenuation of echoes from deeper structures within the tissue and non-linear amplitude compression to allow the display of both the high amplitude specular reflected echoes and the much lower amplitude returns from the distributed scatterers within the tissue on the limited dynamic range of the display(ref.6). Time differentiation is sometimes employed to accentuate small differences in echo amplitude. It is found that the optimum signal processing characteristics vary markedly as different areas and organs are examined.

1.4 Properties of medium

Sonar and radar operate in a medium which is relatively homogeneous and seek to detect and locate a number of discrete targets. For range calculation purposes the velocity of propagation can be regarded as a constant. The targets can be regarded as omnidirectional reflectors, that is, each target will reflect energy in many directions and the relative angle between the target and the beam is not a critical factor.

In sonar, the medium causes problems due to vertical velocity variations, causing ducting in the medium and creating dead zones into which the sound does not propagate. In radar there are problems due to attenuation in the medium from rain and cloud, and both sonar and radar suffer from clutter problems due to reflections from the boundaries of the medium.

Exploration seismology operates in an inhomogeneous layered medium and seeks to delineate a continuum of targets representing the boundaries

between layers in the medium. The targets are large in extent and reflection is specular. Problems due to the medium include overlapping of echoes from successive layers and the presence of non-layered structures such as faults, salt domes and so on. Reverberation or multiple reflection within layers gives rise to echo returns at increased apparent depth where no actual interface exists. This effect is particularly serious in marine exploration where reverberation in the water layer provides much larger signals than the reflection from geological structures below. The velocity of propagation varies widely from 1.5 km/s in water to 5 km/s in hard rock at depth.

In ultrasonic pulse-echo visualisation, the medium consists of inhomogeneous non-layered structures and useful information is contained in large specular reflections from the surfaces of organs and structures and small scattered reflections from small inhomogeneities within the structures. Coherent noise such as reverberation or multiple reflection within the tissue does exist and is more of a problem than random noise. The velocity of propagation is relatively constant, lying in the range 1.4 - 1.65 km/s, and variations in velocity carry useful information.

Local variations in attenuation occur and may be observed by their effects on the echoes from underlying tissue. These effects are most clearly seen in simple scans, that is, when there is no overlap between separate lines of sight. Local areas of reduced attenuation, found in cysts, cause an enhancement of underlying tissue. Local areas of increased attenuation in bone, some carcinomas and foreign bodies, cast an ultrasonic shadow (figure 2).

2. LIMITATIONS OF PULSE ECHO ULTRASOUND

2.1 Beam resolution

The design of a transducer for pulse echo visualisation is a compromise between azimuthal resolution and depth of field. A plane transducer has a beam which is cylindrical in the near field and of the same size as the transducer aperture and extends into a cone in the far field. Far field beam width can be reduced by increasing the aperture but this increases the beam width in the near field. An improvement in performance can be obtained by focusing and the larger the aperture the better the resolution at the focus but the less the depth of field and the worse the resolution outside the focal area (figure 3). The beam width employed in the examination of the abdomen and pregnant uterus is in the order of 10 to 20 mm over a range of 200 mm. This is considerably worse than the axial resolution which can easily be made of the order of 2 to 3 mm and can be reduced below 2 mm at the price of transducer sensitivity. Thus the resolution cell on the echogram is an ellipse with the major axis the lateral or azimuthal resolution and the minor axis the axial resolution (figure 4). The orientation of the resolution cell is variable throughout the echogram, depending on the directions from which each part of the field is viewed. This precludes deblurring or resolution improvement by two-dimensional picture processing on the completed echogram, as the blur function is neither shift invariant nor even determinate at a particular point.

2.2 Imaging rate

The imaging rate for pulse echo visualisation is set by the time required to receive echoes from a single line of sight and the number of lines of sight needed to form the echogram. In the pregnant uterus, the required depth of 250 mm sets a time delay of 330 μs and approximately 2500 lines are needed to make a high quality compound scan echogram. Therefore the theoretical time required to obtain the data is of the order of 1 s.

Normally, additional time is used between pulses to allow echoes to die away and to allow for the water path transit time in water coupling scanning. With current machines, compound scan echograms are obtained in about 15 s. Reducing the number of lines of sight tends to reduce the detail and clarity in the echogram but also provides a reduction in acquisition time. A number of instruments have appeared on the market which use approximately 50 lines of sight with frames repeated up to 50 times/s to provide real time viewing, for instance for the beating heart. The picture quality in this technique is considerably reduced but the ability to view dynamic structures in real time provides another dimension in ultrasonic diagnosis.

2.3 Artifacts

The main artifacts in ultrasonic pulse visualisation are due to multiple reflection of the ultrasonic pulse within the propagation medium(ref.7). If the pulse is reflected back and forth along the beam axis between two reflecting surfaces before returning, it arrives at the transducer at a time delay which does not correspond to the distance to a physical interface. The largest multiple reflection is that which twice traverses the distance between a reflector and transducer face giving rise to an apparent structure which is exactly twice as far from the transducer as the real interface. The presence of this artifact makes it necessary in water coupling scanning to have the transducer as far from the skin surface as the greatest depth of penetration required in the patient. This ensures that the twice round echo from the transducer face to the skin surface arrives after the last echo of interest from within the tissue. In skin contact scanning, these artifacts occur within the examined field (figure 5(a), (b)). The most obvious occurence is when there is a large reflector such as an air containing structure or a bone close to the surface and up to 10 to 15 artifact echoes may be seen. This type of artifact also occurs due to multiple reflection within the layers of the anterior abdominal wall giving rise to one pattern consisting of strong signals from the genuine echoes and a weaker set of echoes from the twice round multiple reflection. This echo is most apparent in the longitudinal scan of the pregnant uterus where a structure may be seen resembling the placenta on the anterior abdominal wall which continues down into the patient's bladder. This type of artifact is strongest in skin contact scanning because the interfaces at the skin surface consist of a skin/transducer material interface which has a large reflection coefficient. In water coupling scanning the corresponding interface is between the skin and the water medium which has a much smaller reflection coefficient and therefore the corresponding artifact is of a much lower amplitude and is rarely seen (figure 5(c)).

In non-axial multiple reflection the multiple propagation path is not on the beam axis, and may not even be in the plane of the echogram. In this case recognition of the echo as an artifact is considerably more difficult and rests mainly on recognition that no reflector should be present in the particular area (figure 5(d)).

3. NEW DEVELOPMENTS AT THE ULTRASONICS INSTITUTE

3.1 Annular array

As mentioned earlier a compromise is needed between depth of field and resolution at the focus. A sharper focus results in a reduction in the depth of field and thus in the usable range of the transducer. One solution to this problem is to construct the transducer with a number of coaxial elements(ref.8). By varying the time delays of the electronic signals associated with these elements a focusing effect may be obtained (figure 6).

Thus in transmission a beam focused at any required distance may be obtained and the focal length altered for consecutive transmitted pulses. With reception the time delays which provide the focusing effect and fix the focal depth may be altered while echoes are being received, so there is effectively sharp focusing on reception throughout the operating range. This approach removes the requirement for a large depth of focal zone and allows a greater aperture and sharper focusing to be used, thus producing a better resolution at each distance. An annular array 130 mm in diameter consisting of a central disk and six annular rings has been constructed at the Ultrasonics Institute to operate at 2 MHz in the abdominal water coupling echoscope. With this system a beam width of 4 to 6 mm has been achieved over an operating range of 180 mm (figure 7).

3.2 U.I. Octoson

The U.I. Octoson is a general purpose multi-transducer water coupling echoscope designed to remove the limitation of scanning time due to the need to move a single transducer over the patient in the formation of a compound scan. It is also designed to be suitable for examination of all parts of the body able to be usefully examined by ultrasound with the possible exception of the eye. The scanner has a scan arm containing eight transducers arranged on the circumference of an arc and pointing towards the centre (figure 8). These eight transducers are mechanically coupled so that they all simultaneously perform a sector scan. Electronic switching is provided so that each transducer is activated in turn to provide an individual line of sight from eight scan positions. The cycle is repeated as the transducers are scanned so that a complete compound scan containing 2400 lines of sight is obtained in about three seconds. The scan arm is supported by a structure with five degrees of freedom to allow translation in three dimensions and rotation and tilting motions. This allows the scan arm to be moved to any position as required to provide a cross section of the body. The whole scan mechanism is immersed in a water tank and the patient lies on a couch above the tank and is coupled either by immersing part of the body in the water or by applying the body to a plastic membrane fitted to the aperture in the top of the tank (figure 9). This system has been found satisfactory for examination of the pregnant uterus, abdominal organs, breast, thyroid and infant brain.

A licensing agreement has been negotiated between the Australian Department of Health and an Australian company, Nucleus Holdings Pty. Ltd. for the manufacture and distribution of this instrument throughout the world.

3.3 Digital signal processing

Present research on digital processing of ultrasonic signals is in its early stages(ref.9). The computer installation consists of an Interdata model 80 processor, a fast 16 bit mini computer, 64 KBytes of MOS semiconductor main memory with 6 MBytes per second memory access rate, two cartridge disk drives with a capacity of 2.5 MBytes per cartridge and 8-channel 20 kHz analogue-to-digital and digital-to-analogue equipment. There is a custom micro-processor interface which controls a 10 MByte 8-bit A/D converter with an 8 channel multiplexer and interfaces with the 6 MByte per second memory access port. It is programmable and can be used for direct input of ultrasonic data or for picture input from a TV signal. It can also be used for XYZ picture output at a rate of 3 points per μs or for TV output. Primarily, investigations in signal processing have centred around the techniques of deconvolution using a least squares spiking filter developed for the processing of seismic data(ref.10). It has been shown that considerable resolution improvement can be achieved both in the axial response and, more importantly, in the azimuthal or beam width resolution.

4. FUTURE DEVELOPMENTS

4.1 Realtime systems

As outlined above, the imaging rate from an ultrasonic system can be increased at the cost of number of lines of sight and therefore picture quality and clarity. In the analysis of dynamic structures within the body rapid imaging rate is essential. There is a variety of approaches being adopted to the provision of realtime ultrasonic systems, ranging from various rotating and sectoring mechanical scanners to transducer array systems. The simplest approach is just to provide a transducer which rocks backwards and forwards in a sectoring motion. However, this is limited by the inertia of the oscillating system. There are two array approaches used: the first consists of a set of 20 or so small transducer elements arranged in a line; each element is used as a separate individual transducer thus giving a picture containing the same number of lines of sight as there are elements. In this simple approach no beam steering or dynamic focusing is employed. In an extension of this approach a transducer with many elements may be used and a group of these elements is used as the transducer to generate a single line of sight. In this way advantage is taken of the large aperture to provide greater resolution. Fixed focusing on transmission and dynamic focusing on reception can be implemented with this system. Another approach is to use an array of about 16 to 20 elements to generate a beam which scans in a sector. Again, focusing in the plane of scan can be achieved. With these electronically scanned systems frame rates to match TV equipment are obtainable.

4.2 Computer processing

Techniques for resolution improvement along and across the ultrasonic beam have been mentioned earlier. Other future developments in the application of computer processing to the ultrasonic data are in the area of artifact removal, tissue characterisation and display. The techniques of dereverberation in seismic data will be investigated in an attempt to achieve axial multiple reflection artifact removal. It is likely that this will be a relatively straight forward procedure when there is a large number of reverberating artifacts as is the case with reflection from bone or air. The problem when there is only a single additional reflection such as in anterior wall artifacts is more difficult. Another approach to artifact removal involves the comparison of data taken from different directions, in a technique which is equivalent to stacking of seismic data.

In tissue characterisation, a number of parameters can be studied. On a one dimensional basis, the frequency response of returned echoes should contain information about the transmission properties of the medium and the reflecting properties of the reflector. If reflection is from a group of scatterers, texture information should be obtainable from the time distribution of echoes and attenuation information from the rate of decay of the amplitude of echoes from an area which may be regarded as composed of uniform scatters. On a two dimensional basis, techniques will be investigated using an approach similar to the velocity analysis and migration stack techniques of exploration seismology.

With the increase in acquisition rate for ultrasonic echo data and the attendant faster turn around of patients in clinical examinations, it is important to stream-line interpretation of the data and the issuance of clinical reports. At this stage ultrasonic data is presented on photographic prints as a set of 20 serial transverse cross sections and 5 serial longitudinal cross sections. The interpreting physician examines this set of data, mentally constructs a three dimensional picture of the scanned anatomy and carries out diagnostic interpretation on the character of echo

returns from areas of particular interest for the diagnosis. It is relatively straight forward in concept to consolidate the data into a three dimensional image of which any required cross section can be displayed(ref.11). It is less straight forward in concept but obviously will be useful to allow the specification of regions of interest and to have automatic analysis of signal character in some quantitative terms such as time frequency response or spatial frequency response.

5. ACKNOWLEDGEMENTS

The paper is published by permission of the Director-General of Health.

REFERENCES

No.	Author	Title
1	Kossoff, G.	Ultrasonic Visualization of the Uterus Breast and Eye by Grey Scale Echography. Proc. Roy. Soc. Med. 67, 2(February 1974); 135-140
2	Urick, R.J.	Principle of Underwater Sound for Engineers. McGraw Hill, 1967
3	Skolnik, M.J.	Introduction to Radar Systems, McGraw Hill, 1962
4	Anstey, N.A.	Seismic Prospecting Instruments, Vol.1, Gebruder Borntrager, Berlin, 1970
5	Robinson, D.E. and Kossoff, G.	The Imaging Properties of Ultrasonic Pulse-echo Visualization systems. Seminar, Optics in Medicine III. Proc. Soc. Photo. Dpt. Inst. Eng. V47, 1974; 84-92
6	Kossoff, G.	Improved Techniques in Ultrasonic Cross Sectional Echoscopy. Ultrasonics, 10 (1972); 221
7	Robinson, D.E., Kossoff, G. and Garrett, W.J.	Artifacts in Ultrasonic Echoscopic Examinations. Ultrasonics 4,4 (1966); 186-193
8	Kossoff, G., Radovanovich, G., Garrett, W.J. and Robinson, D.E.	Annular Phased Arrays in Ultrasonic Obstetrical Examinations. Ultrasonic Diagnostics. Plenum Press 1975 (In press)
9	Robinson, D.E., Williams, B.G. and Horn, P.	An interactive acquisition and processing system for ultrasonic pulse echo signals. Ultrasound in Medicine and Biology (In press)
10	Robinson, E.A.	Multi Channel Time Series Analysis with Digital Computer Programs. Holden-Day 1967
11	Robinson, D.E.	Display of Three Dimensional Ultrasonic Data for Diagnosis. J. Acoust. Soc. Amer. 52, 2(Pt.2) (August 1972); 673-687

TABLE 1. COMPARISON OF BEAM CHARACTERISTICS OF VARIOUS PULSE-ECHO AND IMAGING SYSTEMS

	Range/Aperture	Aperture/Wavelength	Range/nearfield length
Aircraft Surveillance Radar	2,500	20	5,000
Airport Surface Detection Equipment	500	400	5
Weather Radar	300,000	6	200,000
Anti submarine Sonar	10,000	3	10,000
Exploration Seismology	10	70	0.5
Water-Coupling Ultrasound	6	100	0.3
Skin contact Ultrasound	10	30	1.5
Terrestrial Optical Telescope	40,000	100,000	1.5
Portrait Camera	100	50,000	0.001

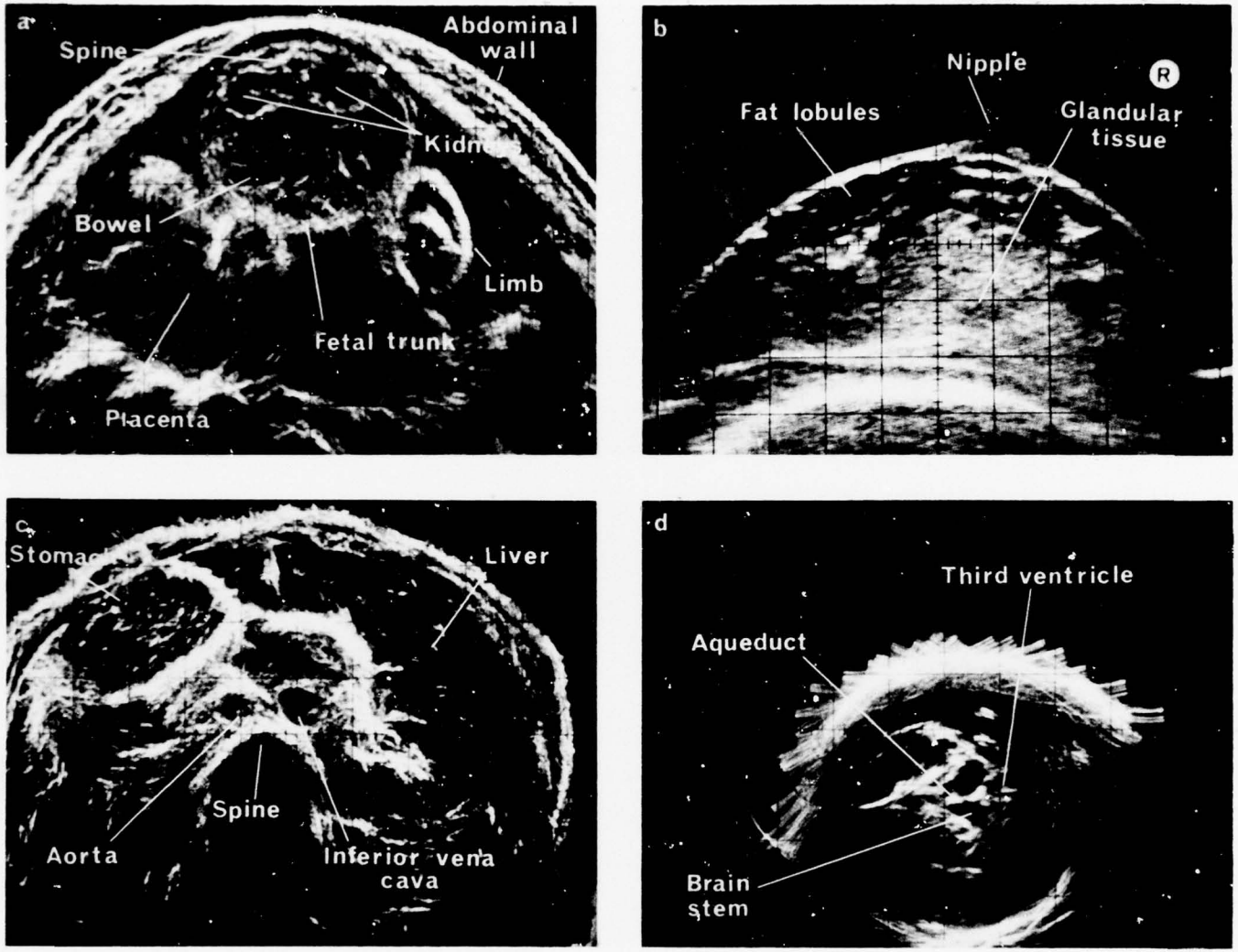
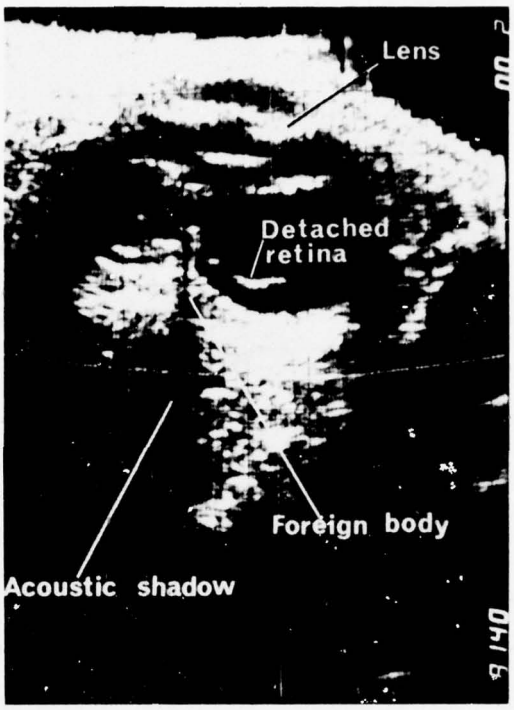
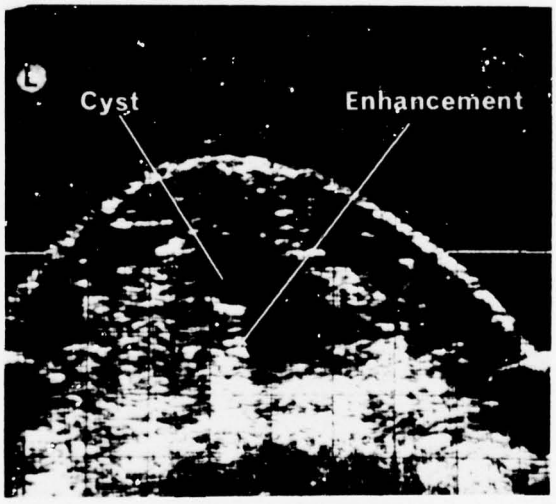


Figure 1. Typical grey scale cross-sectional ultrasonic echograms of normal anatomy, (a) transverse section of the pregnant uterus and fetal trunk in late pregnancy; (b) normal pre-menopausal breast; (c) transverse section of the trunk showing liver, kidney and great vessels; (d) transverse section of infant brain



(a)



(b)

Figure 2. Illustration of the effect of local variation in attenuation;
(a) an increase in attenuation or reflection gives rise to shadowing or observing of echoes from underlying structures;
(b) a decrease in attenuation produces a region of enhanced echoes from deeper echoes.

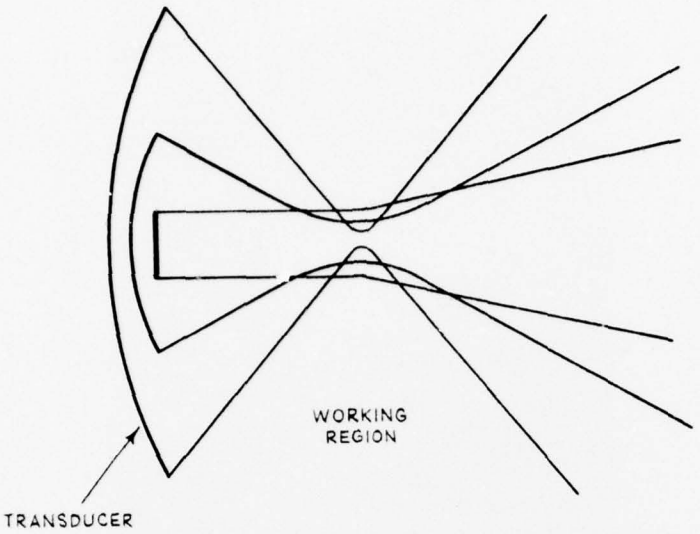


Figure 3. Effect of transducer aperture and focus on the beam pattern. A large aperture produces sharp focus with better resolution within the focal region but worse resolution elsewhere

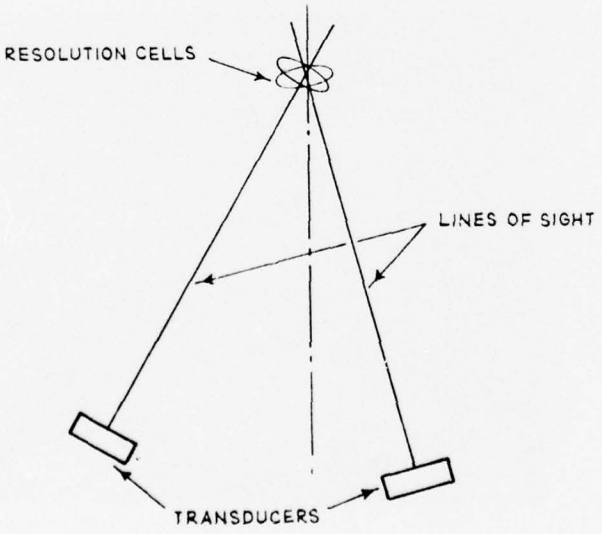


Figure 4. Illustration of the resolution cell in the cross-sectional visualization process. The axial resolution is normally several times better than the azimuthal. The orientation of the elliptical cell is dependent on the direction of the line of sight and varies throughout the echogram.

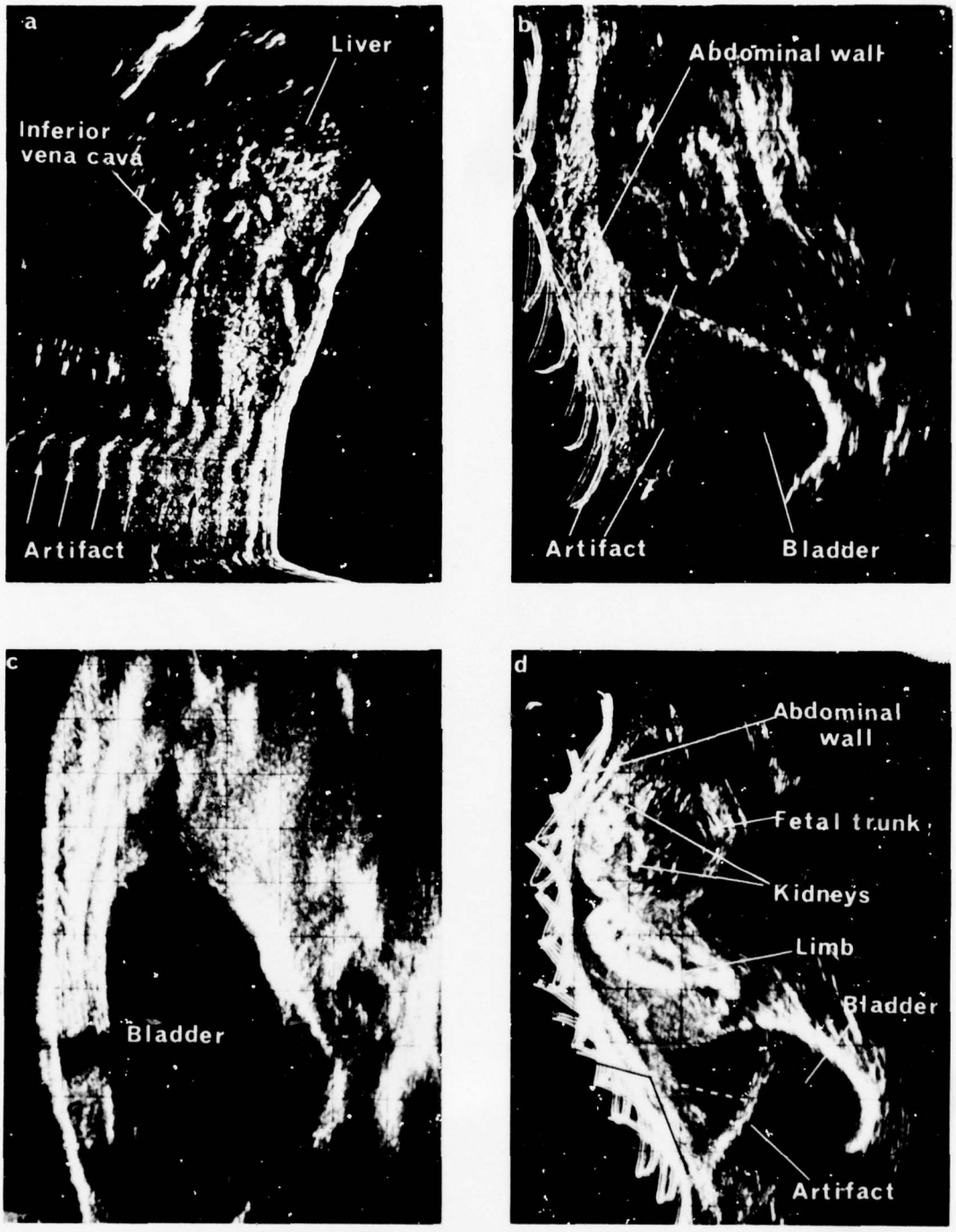


Figure 5. Multiple reflection artifacts give rise to a number of characteristic appearances; (a) "ring down" from reflection between the transducer face and an air containing loop of bowel in a contact scan giving rise to at least 8 discrete artifactual echoes; (b) multiple reflection between the transducer face and layers within the abdominal wall giving rise to apparent "placental" echo structures with the bladder; (c) a similar situation with water coupling showing the absence of multiple reflection artifacts due to the reduced reflection from the skin/water interface compared with skin/transducer; (d) a non-axial multiple reflection artifact. The beam was reflected at the anterior surface of the bladder and propagated in the abdominal wall to a strong reflector, probably the skin/air interface or the pubic bone, and thence back along the same path to the transducer.

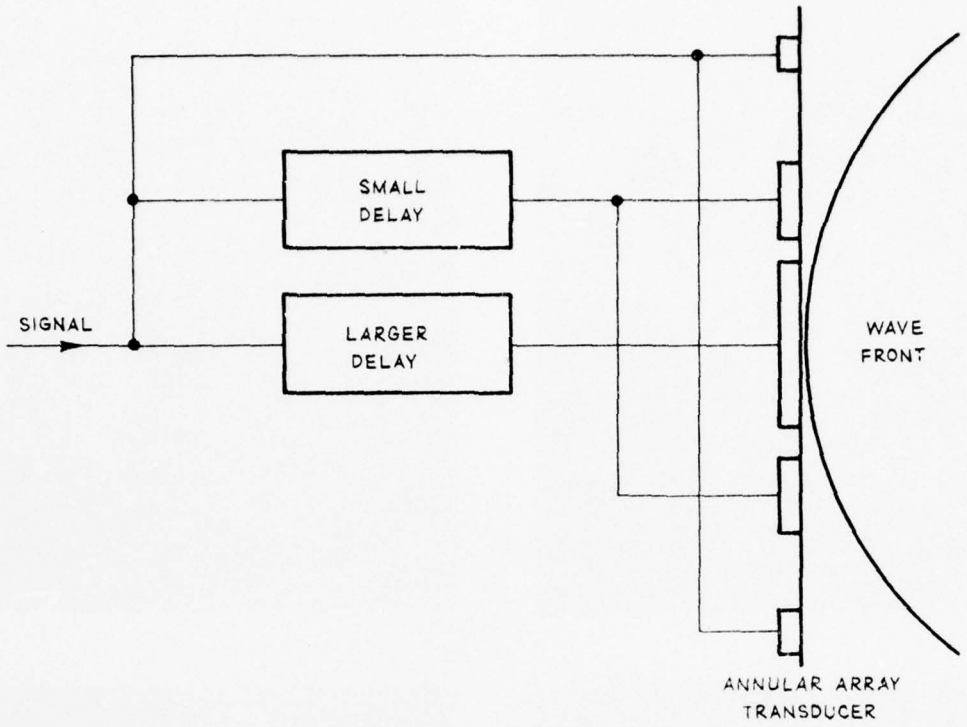


Figure 6. Arrangement of annular array transducer. On transmission the outer rings are excited first and the central disc after a delay period to form the wave-front. On reception the signals from the central disc must be delayed until the contribution from the outer rings are received and then combined to form the electrical signal.

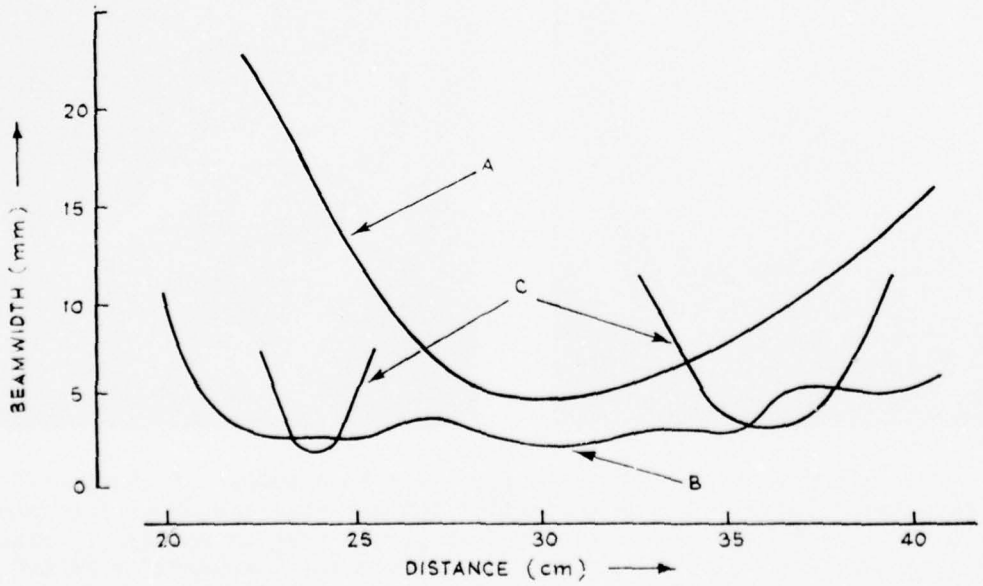


Figure 7. Effect of focusing on focal zone. Curve A is the pattern of a fixed focus 7 cm aperture which is optimum for fixed focus at 2 MHz over the range 20 to 40 cm. Curve B is the beam pattern obtained experimentally using the annular array. Curves C are two calculated diffraction-limited patterns for fixed focus using an aperture equal to that of the annular array.

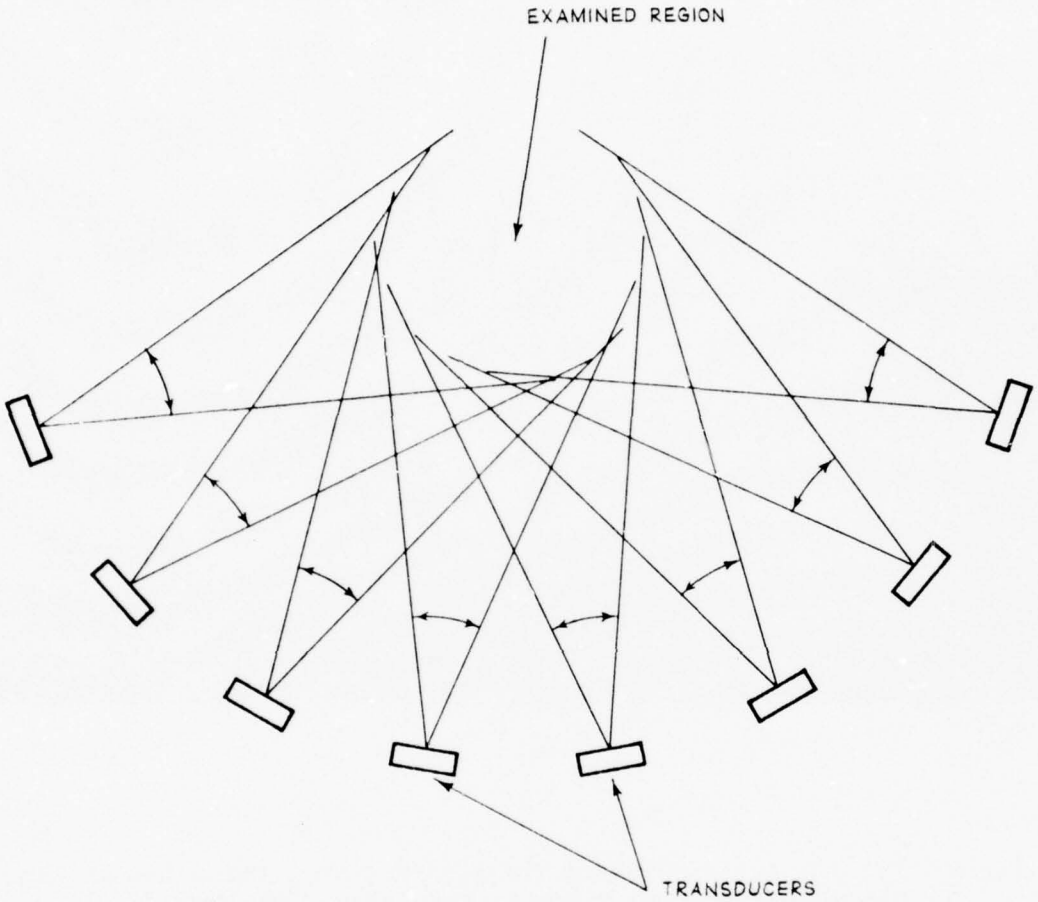


Figure 8. Scheme of operation of the U.I. Octoson. All transducers are mechanically coupled and sector scan simultaneously. Electronic switching is used to cycle rapidly around all eight transducers so that effectively eight lines of sight are obtained for each increment in the sector scan. In this way a complete picture can be obtained in two to three seconds.

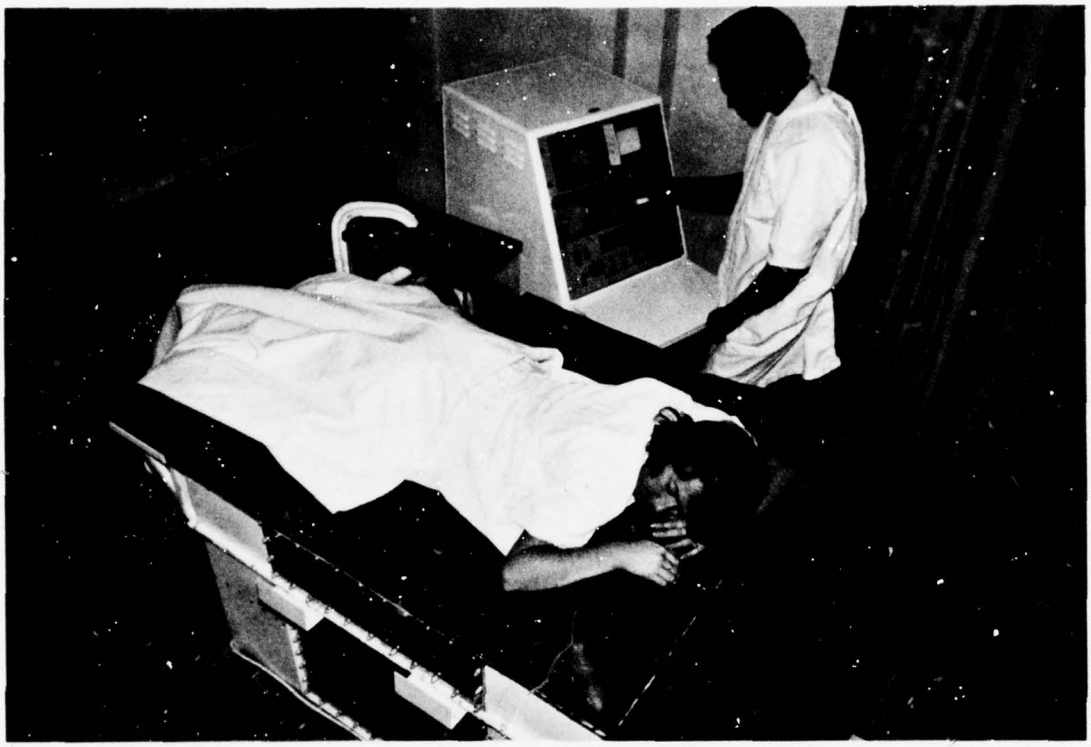


Figure 9. The U.I. Octoson showing the general layout and method of coupling the ultrasound between the transducers and the patient.

SIGNAL PROCESSING IN RADIO ASTRONOMY

P.J. Wild

Chief, Division of Radiophysics, CSIRO

1. INTRODUCTION

I suppose astronomy is the oldest of the sciences. Astronomical observations have in fact been made for thousands of years using the unaided eye and for several centuries using optical telescopes. On the other hand, radio astronomy is only 30 or 40 years old but, in spite of its very brief history, it has probably already overtaken optical astronomy in the art of signal processing. The reason is perhaps that the optical people have had too easy a task; they had two ready instruments - the human eye and the lens - which produce a focused image with an aperture of thousands of wavelengths, and until quite recently complacency and conservatism marked their attitudes to instrumentation.

Radio astronomers have had much tougher problems to solve. As an example, to obtain a 30 second-of-arc angular resolution, one needs a 6000 wavelength aperture. You can achieve this optically with a 3 mm aperture. In radio astronomy, at 21 cm (the most important wavelength), you need 1.2 km to do the job. So, when you are faced with the need for this kind of resolution, you find yourself outside the realm where a single aperture can be used.

Radio astronomers have had to live on their wits and in the process have acquired perhaps a deeper understanding of the underlying principles involved. At one time - not many years ago - 1° was considered a very creditable resolution at radio frequencies. Nowadays the angular resolution is approaching - and in some cases surpassing - that obtained with optical telescopes. Figure 1 gives an example of a modern radio picture of an external galaxy (M101). This was taken with the aperture synthesis telescope at Westerbork (the Netherlands), which is the leading instrument of this type in the world today. The angular diameter of the nebula is 30 min of arc. The wavelength, 21 cm, is that of the ground-state hydrogen line, so the picture gives the distribution of hydrogen within the galaxy. One makes use of the Doppler effect to show the relative velocities of the different parts of the galaxy. The red is, relatively speaking, receding, and the blue approaching, so you can see that the whole picture is consistent with a whirling spiral-arm galaxy which is rotating about its centre in a plane tilted with respect to the observer.

The angular resolution in figure 1 is 30 second-of-arc. One can do better than that with this instrument by going to shorter wavelengths. Within five years, radio astronomy will be attaining a fraction of a second-of-arc resolution - better than the best optical telescopes, which are limited by the atmosphere. Figure 1 is in fact a triumph not so much of optics, but of signal processing.

2. IMAGE FORMATION AND APERTURE SYNTHESIS

In talking of signal processing I shall be dealing mainly with this central problem - of how to use signal processing as an aid to achieving images at high angular resolution. Our aim, then, is to register the brightness distribution of celestial objects in the sky as a function of direction with the maximum possible angular resolution determined by the aperture of the instrument in wavelengths. As a starting point, consider figure 2(a) which illustrates how this is done in optics. We use a lens to bring the radiation to a focus. If we want to do the same in radio, one very direct method (figure 2(b)) is to use an array of elements and join them to different points in the image plane through lengths of cables which obey precisely the same rules as the optical model. In this way we can use an array of elements to produce a simultaneous image using what is called a branching network.

Figure 3 shows virtually a reproduction of the array of figure 2(a), in which the excess path lengths required to focus rays from the θ direction are seen to be given by $x \sin \theta$. We can write down the expression for the power received, P_θ, which, when expressed as a double sum, can be seen to depend on the cross products $\langle v_n v_m \rangle$ of the voltages v_1, $v_2 \ldots v_n \ldots v_m \ldots$ induced in the elements of

the array. The distribution of power, P_θ, in the image is just the Fourier transform of these cross products $\langle v_n v_m \rangle$. An important point is that the $\langle v_n v_m \rangle$'s are time averages and do not have to be measured at the same time; they do not even have to be measured at the same position on the Earth as long as the spacing and orientation are correct. Since we are dealing with the cross products between pairs of spaced elements, we can simulate, for instance, a 7-element array with only 4 elements (see figure 4); we obtain our image by using six correlators and taking the Fourier transform of their outputs.

We can extend this principle to two dimensions, in which case we need to measure every correlation at every spacing and orientation over a plane, which we call the aperture plane. Once again we can economize very greatly in the number of elements we use; we do not have to move aerials all over the ground to acquire the 2-dimensional pattern. The most practical way is to use an east-west array of movable elements in which we take a number of correlations simultaneously and use the rotation of the Earth to vary their orientation. That is the technique of aperture synthesis.

The history of aperture synthesis is interesting because it was first proposed by Dr. J.L. Pawsey in 1946 in Australia and the first person to use earth rotation was also an Australian - Dr. W.N. Christiansen. However, the real development in aperture synthesis was done by Sir Martin Ryle at Cambridge who, with splendid persistence, greatly aided at a vital stage by the introduction of electronic computers, perfected the technique. He won himself a well-earned Nobel prize for the job which he did. What aperture synthesis has done is to change radio astronomy from minute-of-arc resolution to second-of-arc resolution and is equivalent to the change in optical astronomy from the unaided eye to a large telescope.

Figure 5 shows a picture of the Westerbork synthesis telescope. It consists of 10 elements about 25 m in diameter, two of which are movable on rails. If my memory is right, they take 16 correlations at any one time and build up a complete picture in some four days. This instrument was responsible for producing the picture in figure 1 and many other fine astronomical pictures.

3. THE CULGOORA RADIOHELIOGRAPH

I now want to consider another aspect of high-resolution radio astronomy: the recording of astronomical phenomena, such as the disturbances on the Sun, which vary rapidly with time. In this case, the aperture synthesis technique employing earth rotation is no longer possible because it is much too slow. Instead one must use an array which records all the different components simultaneously and then somehow perform a Fourier transform, preferably in real time; in the typical case of the Sun one would like pictures at intervals of about one per second.

Now there are several known ways of performing the Fourier transform. One that we have already seen is to use a branching network which is a fairly involved piece of hardware when the number of image points is large; another is to use simple scanning techniques but these are very wasteful in sensitivity. Another is to use a correlator and the Fourier transform method as in aperture synthesis but that makes very heavy demands on the capacity of the computer because the process has to be done so quickly. Or one can use analogue methods such as described in the paper by Dr. Cole at this Symposium.

I would like to describe one instrument that solves this problem in a unique way: the instrument is the CSIRO Culgoora Radioheliograph, which has now been operating for some seven years and is in fact still the only fast image-forming radio telescope in existence. In designing this instrument, it was first necessary to choose an aperture configuration; we chose an annular array because we were very keen on having circular symmetry so as to avoid major sidelobe distortion in certain directions. Figure 6 shows an artist's impression of the

instrument. It consists of 96 dishes, each 13 m in diameter, arranged in a circle of diameter 3 km. All the dishes act coherently, giving, from the point of view of resolution, an aperture 3 km across. Initially the instrument operated at 80 MHz; it is now operating on three frequencies: 40, 80 and 160 MHz and will soon go to 320 MHz.

Figure 7(a) shows the polar diagram of the array with all the elements connected in phase with one another. Within the outer ring (the first-order grating response) the power polar diagram is in fact the function of J_o^2. The grating response results because of the finite distance between the elements. If one had a continuous ring, these responses would not be present. There is nothing one can do about getting rid of grating responses short of filling the ring; what you do is to choose your parameters so that, when you scan the Sun, you make sure that you have enough angular separation (say 2^o) between the centre of the pattern and the first-order grating response.

The main problem with such an array is how to process the signals so that sidelobes of the J_o^2 pattern are suppressed. A somewhat simplified explanation can be given in terms of figure 7(b) which shows the result of connecting all the aerials together with a progressive phase shift all the way round so that when you come round again to where you started you have an exact number of cycles. The polar diagram has a black spot in the middle instead of a white spot; out from the centre, the sidelobes rapidly become very similar to the ones in figure 7(a). What we do is to record each picture point using the polar diagram of figure 7(a) for a few milliseconds and then switch over to that of figure 7(b) for a few milliseconds and subtract one from the other, to obtain the polar diagram shown in figure 7(c) which is just a single dot. Mathematically one can develop the theory of this technique so as to obtain any shaped polar diagram function you wish within the resolution limit of the aperture.

The radioheliograph images contain 48 x 60 picture points. The 48 N-S sets of points are obtained simultaneously using a branching network of the type indicated in figure 2(b) while the 60 E-W sets of points are obtained by scanning.

Figure 8 gives examples of 80 MHz pictures taken with the radioheliograph at the time of a solar outburst, in which a giant magnetic arch prominence erupts outside the optical limb of the Sun (shown as white circle). The pictures span about one hour of activity. Each picture results from superimposing two separate pictures, taken in opposite senses of circular polarization and passed through different colour filters (red and blue/green). The colours therefore indicate the magnetic field polarity of the origin of the radiations.

4. CONCLUSION

Signal processing is also used in various other applications for radio astronomy - for instance Fourier spectroscopy, de-dispersion of pulsar signals, and a wide range of sophisticated techniques (sometimes so sophisticated that they arouse the suspicion of the simple purist) for 'cleaning' pictures to enhance picture quality, and even for attempting to improve resolution beyond the Fourier limit. Each of these topics would require a lecture on its own.

In this paper I have concentrated on the application of signal processing in the quest for high resolution in radio astronomy. I hope it conveys the message that signal processing and computer technology have played an essential role in solving the problems that once seemed insurmountable.

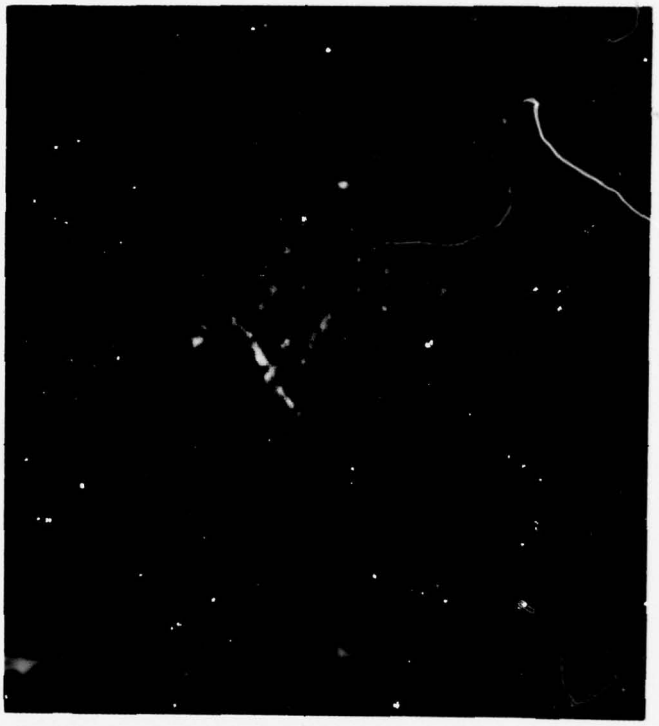


Figure 1. Radio picture showing the distribution of neutral hydrogen in the spiral galaxy M101, made with the Westerbork Synthesis Telescope

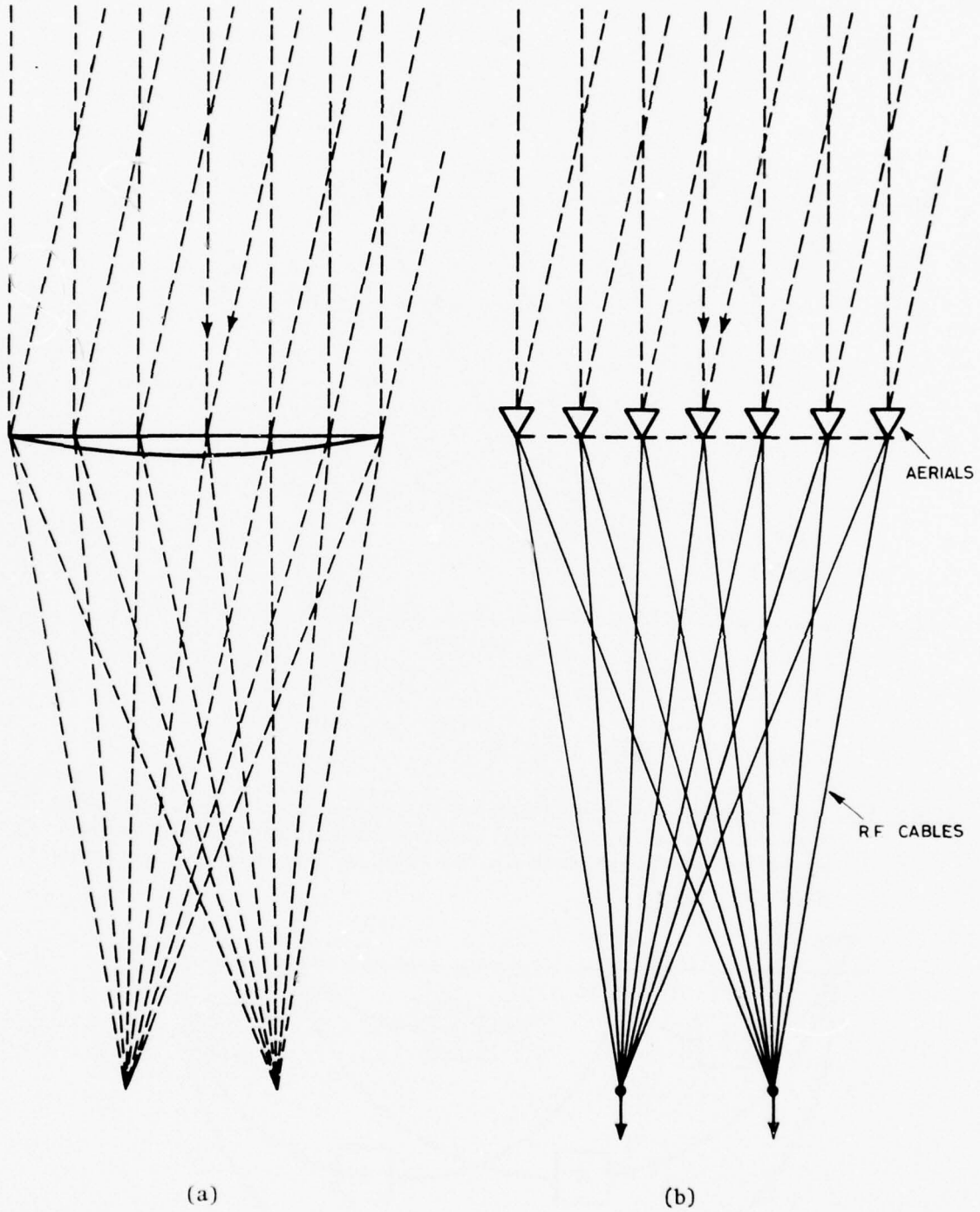


Figure 2. Showing the similarity of principle between a lens (a) in optics and a branching network (b) in radio astronomy

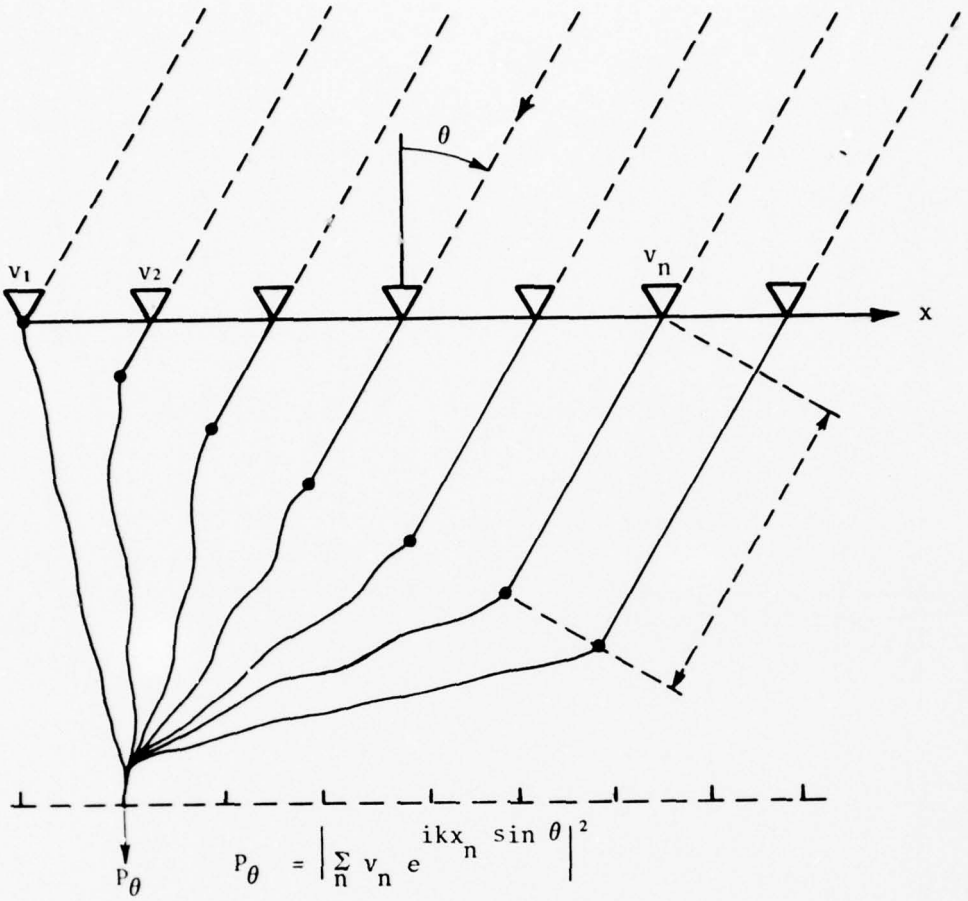


Figure 3. The figure shows connections of a branching network associated with one direction (θ only) and indicates that the brightness distribution P_θ is given by the Fourier transform of the cross products $\langle v_n v_m \rangle$

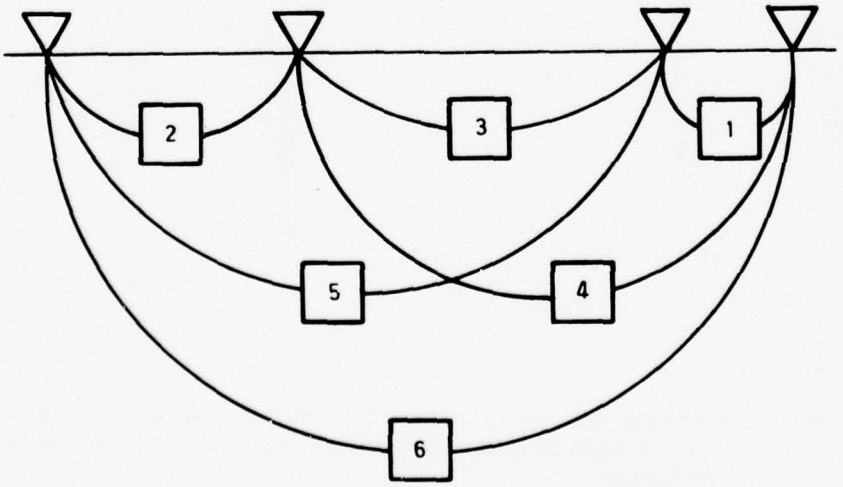


Figure 4. Illustrating how an equi-spaced 7-element array can be simulated by the use of only 4 elements with the aid of 6 correlators

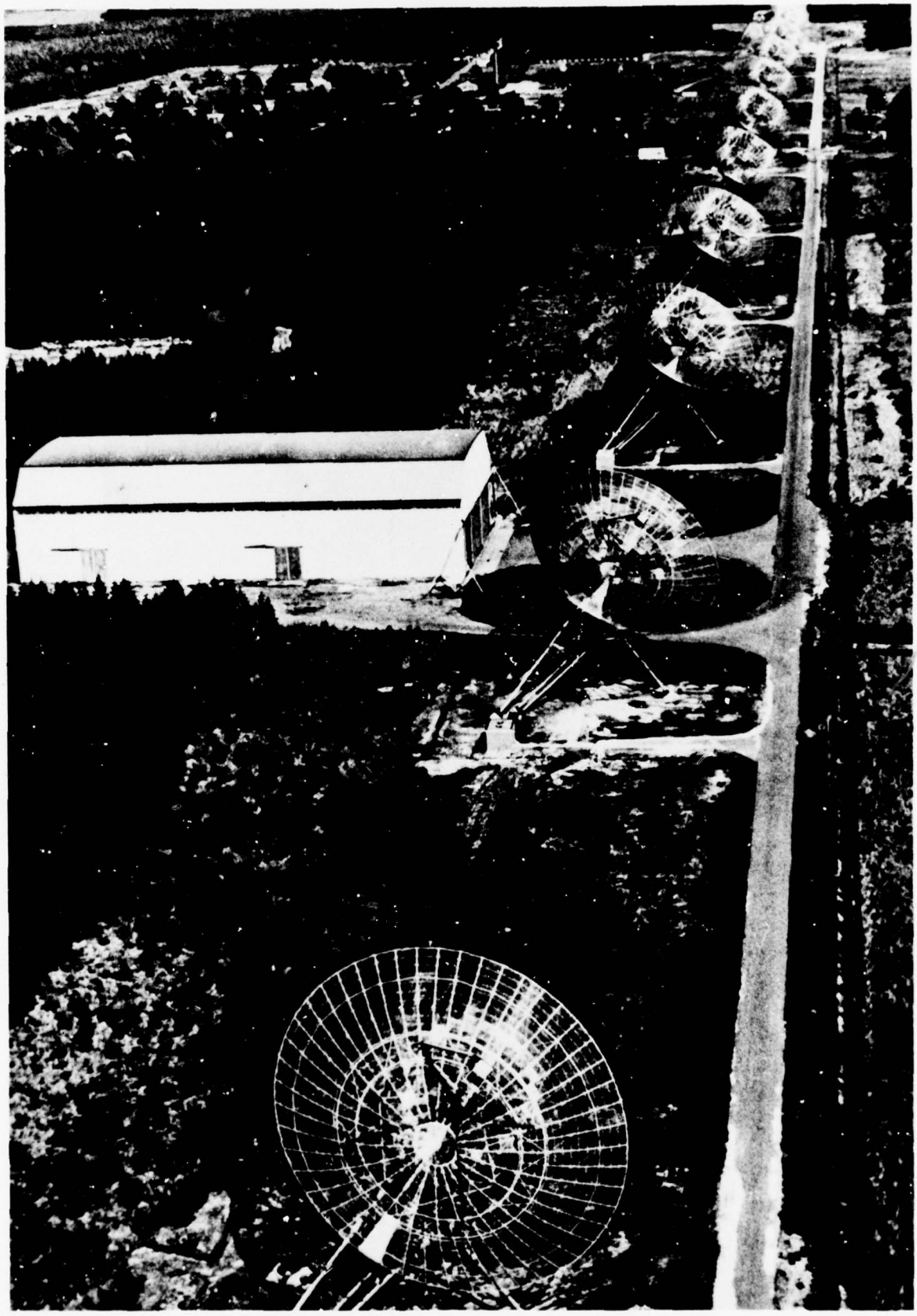


Figure 5. The Westerbork Synthesis Telescope in the Netherlands

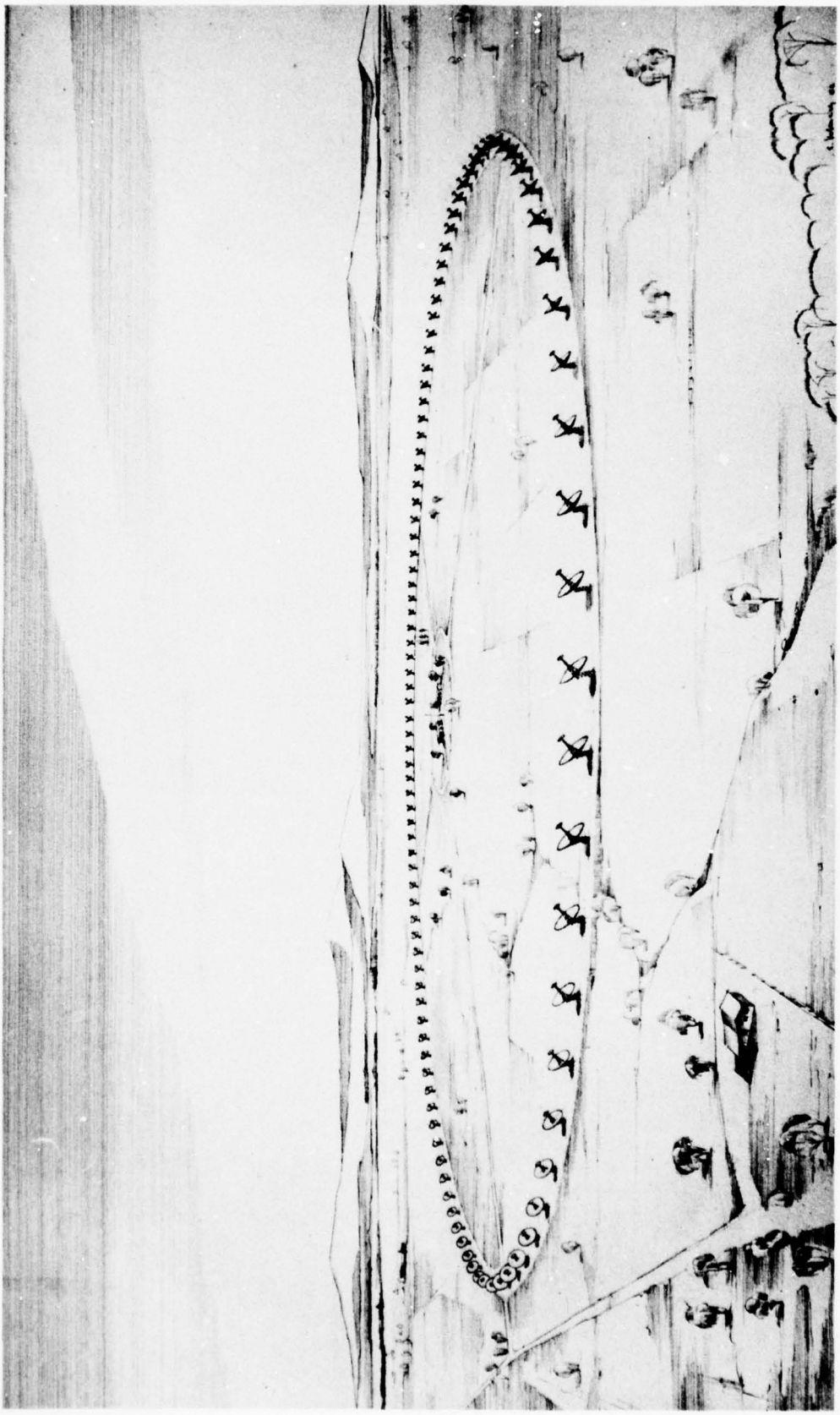
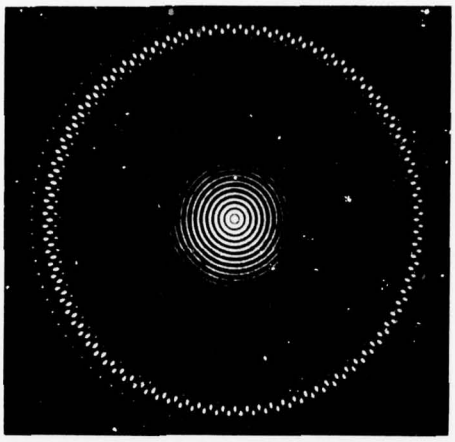
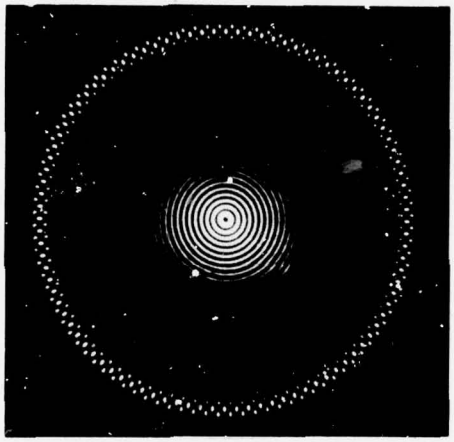


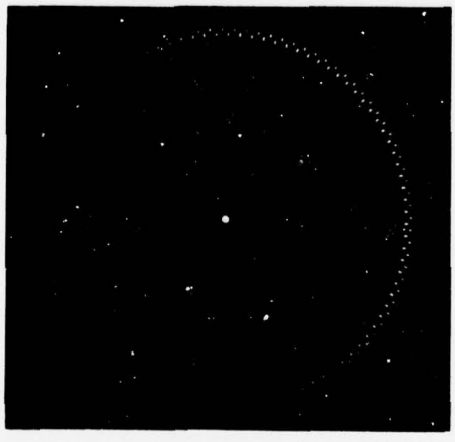
Figure 6. Artist's impression of the Culgoora Radioheliograph



(a)

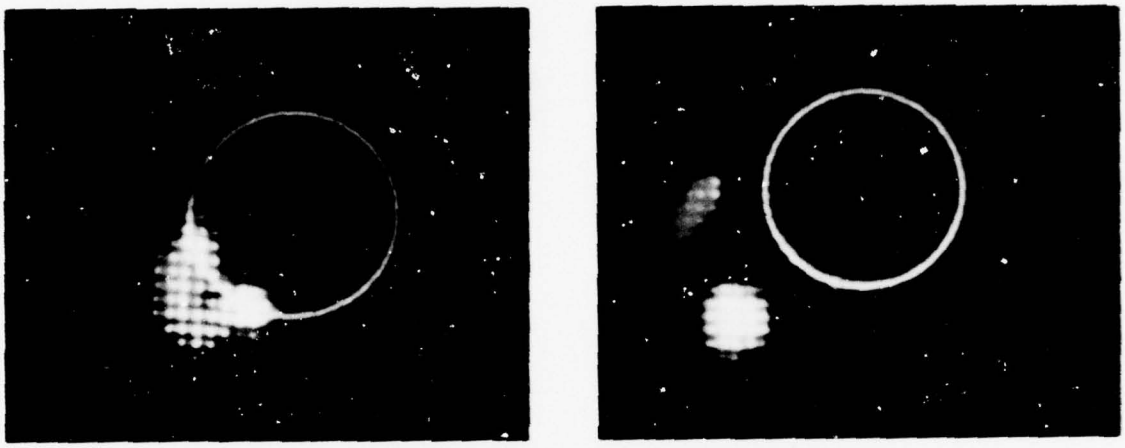


(b)



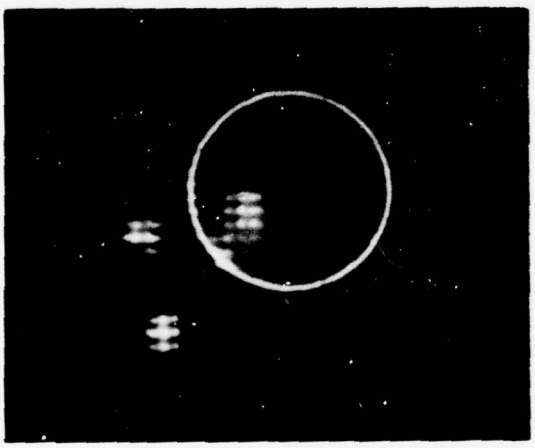
(c)

Figure 7. Polar patterns of the Culgoora Radioheliograph showing how a peincil beam is formed (c) by subtracting pattern (b) from pattern (a). The photographs were obtained using optical modelling of the array

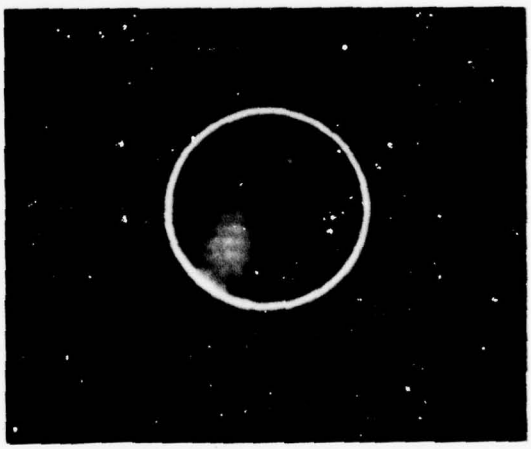


1

2



3



4

Figure 8. Four 80 MHz Radioheliograph pictures taken during the course of a solar outburst. The circles indicate the optical disk of the Sun and the colours indicate polarization as described in the text

ACOUSTIC AND OPTICAL PROCESSING FOR RADIO ASTRONOMY ARRAYS

T.W. Cole, Division of Radiophysics, CSIRO, Sydney, Australia

SUMMARY

The paper discusses optical and acoustic analogue approaches to array processing in radio astronomy. In particular it will be shown how a hybrid approach can be employed to utilize the best properties of analogue and digital methods.

A theme which is developed is that of integration of the analogue device. By taking a systems approach to a processor, it is shown how device simplification and ruggedization can occur.

The applications which are discussed are in radio astronomy. They include acousto-optic radio spectroscopy, broad-band cross-correlation, multi-channel spectroscopy, antenna array beam formation and microwave acoustics.

1. INTRODUCTION

The multidisciplinary nature of this symposium emphasizes the fact that the same concepts of arrays and of imaging can be applied to a large number of situations. Identical mathematics can be used to describe systems employing any type of wave which can be represented by a scalar function of either space or time, for example light, radio, bulk acoustic (shear and longitudinal), and surface acoustic waves. Advantage can be taken of the analogy between wave types to use one type of wave in processing the signals associated with some other type of wave, thus gaining access to special technologies.

This paper explores this concept of transforming one type of wave into another and processing signals in an analogue system. One can attempt to utilize the best features of each of the wave types.

The illustrations to this approach come from radio astronomy and include imaging arrays and radio spectroscopy.

2. RADIO FREQUENCY SYSTEMS

To begin with, a brief review of radio imaging arrays and radio spectroscopy reveals a number of areas where it is expensive and cumbersome, if not impossible, to process in radio frequency waveforms.

Consider first the characteristics of an array type of imaging radio telescope. Such an instrument samples the radio wave field across some aperture (usually plane) with a number of smaller antenna elements: the sampling is usually a thinned or dilute sampling. The signals from all sampling points are combined and processed to form an image of the radio brightness distribution across a region of sky. The signals themselves are broadband noise and the frequency spectral range accepted by the system is defined by filters in the receivers. Hence the system is a partially coherent one and care must therefore be taken that signals are combined with differential path length differences or delays less than the coherence length of the narrowband signal.

Perhaps the major limitation in the radio system is that a radio receiver is only one-dimensional with a single input and a single output. Each receiver channel can measure information about only one point in the image. Simultaneous image formation requires a bank of receiver channels, one for each point in the image plane. The radio frequency region has no simple analogue to the human eye or the photographic plate.

Even assuming that one has available a large number of receivers, it is still not a simple matter to form an image. One method is to measure the relative coherence across the receiving aperture by cross-correlating the signals received at the individual antennas. The image (smeared by some instrumental response function) is then the Fourier transform of this complex coherence function and the transformation is normally performed digitally.

However, for real-time imaging the normal method is to use the radio network equivalent of a lens, known also as a branching network. This consists, in general, of sets of cables joining the individual antennas to summation points with cable lengths such that waves from given directions arrive at the corresponding summation points in phase. In addition to this phase requirement the coherence length requirement must also be satisfied.

A specific example is the Culgoora radioheliograph(ref.1), which has 96 elemental antennas and which forms an image consisting of points on a

48 x 60 matrix. The instrument has 48 receiver channels and a branching network forms the 48 image points in one column of the final image. The complete image of 60 columns is obtained by scanning. But the "network lens" to give even 48 output points contains 96 x 48 = 4608 accurately phased cables. A "network lens" for the whole 48 x 60 point image, with close to 300,000 cables, is clearly impractical.

Almost identical observations can be made about the field of radio spectroscopy where the "image" consists of a number of points along the frequency axis. A multichannel filter system has a bank of filters tuned to adjacent bands of frequencies as the dispersive system and a bank of detectors and integrators as the output channels. Alternatively one measures an auto-correlation function and performs a Fourier transformation. We have then a dual of the imaging case above - a function of time rather than of a spatial dimension.

In this simplistic review of radio astronomy instrumentation many of the practical problems have been ignored but nevertheless four areas can be identified where processing by an analogue wave type might be advantageous. These are:

(i) The cumbersome nature of the "radio network" lens.

(ii) In multiple output systems, the output radio receivers and detector units are complex and expensive.

(iii) In cross- and auto-correlation systems the final Fourier transformation can be both expensive and slow.

(iv) With large antenna systems the variable delays required to maintain coherence can be up to tens of microseconds and these are not only bulky and expensive in radio waveform but can also be unstable.

In the following sections two types of analogue system are described. The first uses coherent optical waves and the second bulk acoustic waves.

3. THE COHERENT OPTICAL PROCESSOR

Consider the coherent optical system shown in figure 1. The beam from a source of coherent light (a laser) is focused through a pinhole and expanded into a collimated beam which illuminates the input plane. Processing is achieved by the Fourier transform relationship which exists between the complex light amplitude distributions in the two focal planes of the main lens(ref.2). The Fourier relationship is quite accurate under the restriction of small diffraction angles. If the amplitude and/or the phase of the collimated light are modified according to some distribution across the input plane, a light detection system in the output plane can record the Fourier transform of the input distribution. An example is shown in figure 1. With spherical lenses, the transform is two-dimensional while for cylindrical lenses, it is one-dimensional. One also notes that in planes other than the output focal plane a Fresnel transform is obtained and this is also illustrated in figure 1.

A Fourier relationship exists between the field distribution in the plane of an antenna and the far-field radiation pattern. Hence by placing an optical analogue of the antenna distribution in the input plane the antenna far-field pattern can be displayed. Extending this one step further, if the signals actually being received by the radio antenna array could modulate the light in the corresponding part of the input plane, the optical output plane would contain an optical image of the radio source distribution being observed (smoothed by the antenna response pattern)(ref.3).

A design study of such an optical processor for the Culgoora radioheliograph was made(ref.4), but no coherent optical processor for radio arrays has yet been put into use. However, the coherent optical processor has been used in radio spectroscopy and the experience obtained offers insight into the concepts of

analogue processing.

The light intensity distribution in the output plane of a coherent optical processor provides a spectrum if a segment of the radio signal modulates the light amplitude and/or phase across the input plane. The problem is to find a simple means of introducing the radio signal into the optical system. One solution is to use the interaction between coherent light and a travelling acoustic wave (ref.5,6,7). The sound wave acts as a travelling phase grating (ref.8) and, when illuminated by a collimated laser beam, produces at least one diffraction order, the angle to which is proportional to the frequency of the acoustic wave. A broadband transducer is used to convert the radio signal to the travelling acoustic wave. The frequency resolution on the spectrum is the inverse of the time it takes for the acoustic wave to cross the optical beam.

In the simplest form of this instrument the spectrum is simply recorded on a photographic film. If the film is continuously moved one obtains a dynamic spectrum, a two-dimensional record of intensity as a function of frequency and time. Such a system is extremely simple and, since all points on the spectrum are formed simultaneously, should be very sensitive when compared with a frequency scanning type of spectrum analyser. Its attractiveness is the simplicity of an optical lens compared with radio frequency filters and the cheapness of the photographic film as a multichannel detector and storage medium.

But this simple system has a number of limitations. One is due to the optical system itself. In any optical system, it is almost impossible to produce lenses to accuracies of better than one-hundredth of a wavelength or so. Even these extremely high-quality figures would give spurious responses or side-lobes at the several per cent level and their random distribution limits the instrument's dynamic range.

A far greater limitation of such a simple system concerns the detector itself. The radio signal of interest is often very weak, superimposed upon a large, constant noise level from the receiver itself or background sources. In such cases photographic film is fogged by this background and the weak signal is lost.

Given these restrictions and the fact that antenna and receiver gain characteristics are usually a function of frequency, the simple acousto-optical spectrograph with direct photographic recording is clearly superior to current approaches only in the situation where the signals of interest are stronger than the background noise and where uniform gain characteristics are not required.

Enormous flexibility is introduced when the photographic film is replaced by an integrating optical sensor that converts the optical image into an electrical signal which can be digitized and fed into an on-line digital computer(ref.9). One such sensor is the self-scanned photodiode array, a linear system of photo-diodes each of which can detect intensity levels with up to a 1000:1 range between saturation level and the thermal noise which the diode array readout system adds to the signal.

A major increase in flexibility comes from the use of a digital computer. The diode array's dynamic range is large enough for the noise levels superimposed on the constant background level to be recorded much better than by a photographic film. The computer removes antenna and receiver bandpass response shapes, applies an intensity calibration and, after removal of the constant background, integrates spectral records to average out the statistical noise. The high accuracy of the digital system can compensate for the fixed imperfections of the analogue components.

These concepts are illustrated in the acousto-optical instruments constructed at CSIRO Division of Radiophysics.

4. ACOUSTO-OPTICAL SPECTROGRAPHS

Two, almost identical, acousto-optical spectrographs are near completion at CSIRO Division of Radiophysics. One is for use at the Culgoora solar observatory to provide sensitive spectra of rapidly varying solar burst phenomena. The other is to be the spectral line backend for the new 4-m-diameter millimetre-wavelength radio telescope being built for CSIRO. Some results are discussed elsewhere (ref.6,7,9).

Both use Bragg effect acousto-optical modulators which accept a 100 MHz wide radio signal and provide a 10 μs interaction between the acoustic beam and the transverse optical beam. The resulting spectrum has a resolution of close to 100 kHz and is recorded by 500-point self-scanned photodiode arrays with a scale of one diode per 200 kHz of spectrum. The photodiode array output is a serial sampling of the integrated spectrum at a standard rate of 20 complete spectral samplings per second. This is digitized and fed into a computer.

In the solar instrument the on-line computer (ALPHA) removes backgrounds and corrects intensities in real time so that the computer provides, via digital-to-analogue converters, an analogue signal representing the calibrated spectrum. At present this is then fed to an oscilloscope trace and photographed but it is also planned to record it on digital magnetic tape for later processing. The resultant spectrum is much more sensitive than the one from the scanning receiver spectrograph while the on-line computer allows accurate calibration of the received intensities.

A spectral line backend for the millimetre-wavelength is required to be very stable, to add little noise to the system and to be able to detect spectral details which are extremely weak with respect to the background noise. To do this, the computer program integrates the spectra and forms a difference between spectra recorded when the antenna is directed on to the source and off the source.

Figure 2 is a photograph of the optical arrangement which is essentially identical for the two spectrographs.

Several improvements are planned for the acousto-optical spectrograph. The performance figures, 500 points across 100 MHz, are already such that it is difficult to achieve similar results with conventional approaches. It is however quite feasible to contemplate acousto-optical modulators with bandwidths of 300 MHz or more(ref.10) and therefore with up to several thousand points. Similarly one can move in the other direction and obtain resolutions of 30 kHz or so over the restricted bandwidth of 20 or 30 MHz. Such a modulator is under construction and will use a ceramic (TLZ-R) shear mode transducer into polished, dense glass (SF56). The advantage here of using the acousto-optical device is then its small size and low cost when compared with other approaches.

A great simplification and improvement is possible when one considers the unit as a complete system. This is illustrated in figure 3. In the first illustration of figure 3 the system is displayed as described above, consisting of the modulator and two lenses. However, it is possible to integrate these three units into one lens(ref.11). That is, one uses the front surface of a lens to collimate the light within the lens. The acoustic beam is in the lens itself and the rear surface of the lens focuses the spectrum to an output plane. The output from the new system, shown in figure 3(b), differs from the more complicated system only in an extra spherical phase term, irrelevant in the situations here where intensity is measured. To increase the illumination efficiency further, the circular laser beam can be converted to an elliptical one, more like the acoustic beam shape, by using an off-axis lens system shown at the bottom of figure 3. By taking this overall view of the system, the acousto-optical spectrograph has been reduced to a laser, beam expander, an output system and a lens whose edge has been ground to accept the acoustic transducer.

The acousto-optical approach to signal processing has several advantages over conventional approaches. It can be used for pure spectral work or it can be used in the cross-correlation of signals from antenna elements used as inter-

ferometers. This latter application offers a large simplification of the radio interferometer system. Two spectra are measured, one with the signals added together in phase, and the other with the signals combined in anti-phase. The difference between these two spectra represents the cross-spectrum, the Fourier transform of the cross-correlation. The system works for differential delays less than the interaction time of the acousto-optical modulator. Since this time can be tens of microseconds, it is therefore possible to use this system for array element separations of several kilometres. The use of the acousto-optical spectrograph combines the conventional requirement for delay compensation cables and a broadband multiplier resulting, once again, in a great simplification.

But for imaging array processing, the coherent optical approach has disadvantages. Being optical, all components must be fabricated and assembled to stringent specifications. Also, the acousto-optical modulator becomes highly wasteful of both laser and radio frequency power. It can take up to several watts of r.f. power to deflect the maximum of several per cent of laser light which one can use in the diffraction order for linear operation. In most systems well over 99% of the laser light is wasted.

Further, in the optical array processor the imaging is done optically where the delay times are negligible. A set of delay cables is still required to maintain coherence in the system. The acousto-optical spectrograph and its applications have significant advantages, but for straightforward array processing it is better to consider purely acoustic processing described in the next section.

5. AN ACOUSTIC IMAGING SYSTEM

An almost complete analogue of the radio system is possible in purely acoustic waves. Such an analogue was suggested by McLean and Wild in 1961(ref.3) and has been used by the Adelaide University group in a low-frequency array used for ionospheric studies(ref.12,13). Basically, piezoelectric transducers are distributed in the same pattern as the elemental radio antennas of the array. Each transducer is excited by the radio signal derived from its corresponding antenna and the resulting ultrasonic waves travel through the medium of the device (for example water) to interfere in a far-field plane. The acoustic intensity across the output plane is identical to the smoothed radio distribution and another array of transducers there can detect the intensity at the points of the image. The output plane of the device is defined by using a large distance between input and output, by using an acoustic lens, or by arranging the transducers on spherical surfaces.

Such an acoustic arrangement is a very much simplified equivalent of the radio branching network and, unlike the acousto-optical array processor, involves little loss of energy. Typically the losses could be the same as for the branching network but without its other defects.

But it is suggested here that, unlike in the optical processors discussed above, the acoustic system can be scaled in delay as well as satisfying the phase requirements of the imaging system. The low sonic velocity when compared with that of light allows a scaling of the radio system such that even for broadband signals the coherence requirements are satisfied and the angles involved in the acoustic device become identical to those of the radio instrument. The processor combines then the branching network and the delay compensation cables. For example, a 3 km radio array involves a radio transmission time size of 10 μs. The delay-scaled acoustic size in water is 15 mm since this is the distance travelled by an acoustic wave in 10 μs.

Given this small size, the question arises as to whether one can arrange an array of transducers on such a small scale. Experimental work has proceeded at CSIRO Division of Radiophysics in the belief that it can and the progress so far is described.

6. AN ACOUSTIC PROCESSOR FOR THE CULGOORA RADIOHELIOGRAPH

Feasibility studies have been undertaken to test the concept of an acoustic processor for the 96-element radio-heliograph. In particular, 48 transducer contacts were established around a ring of diameter 12 mm on the surface of a single slab of piezoelectric ceramic (TLZ-R). The ceramic transducer was resonant in the thickness mode at a frequency of 7 MHz, corresponding to the intermediate frequency of the heliograph. The individual transducers were defined by the position and size of dots of metal which constituted the transducer contact. The actual dots were circular of diameter 0.5 mm. Figure 4 shows the partially connected transducer array as well as the printed circuit board used to connect to coaxial cables. Also shown in figure 4 is the output array (to be mounted at the output plane) consisting again of 48 individual elements but this time arranged as an array of 8 x 6 points of the output image.

The fabrication techniques were the standard ones of microcircuitry. A mask at a scale size of 20 times was carefully cut on a coordinatograph and photographically reduced. The ceramic transducer was vacuum coated with copper and photoresist was spun on. Contact printing and development of the photoresist allowed one to etch away the unwanted copper and leave the desired electrode pattern. In these first experiments contacts were made with conducting epoxy and 25 μm wire. The ceramic was then heated and poled.

In the second version a modified procedure is to be used. The developed photoresist is to be used as a mask in a gold-plating bath so that the resultant electrode pattern is in gold-plated copper. Again with a modified mounting system, it is then possible to use a standard ultrasonic wire bonding machine to contact the individual elements.

For the material and electrode sizes used, the radiation resistance of the transducers when transmitting into water was between 50 and 100 Ω and matching to 50 Ω was achieved by parallel tuning of the transducer capacitance. The bandwidths varied somewhat owing to the conducting epoxy but were approximately 0.5 MHz. This can be increased by using a mercury medium, by backing the transducers with epoxy or by the use of a quarter-wavelength acoustic impedance-matching section. This would slightly complicate the impedance-matching circuitry.

The individual electrodes correspond in size to 3 or 4 acoustic wavelengths at 7 MHz with a separation between electrodes of at least a wavelength. The electrodes behave almost independently with some interaction still to be accurately measured.

The microcircuitry techniques can produce patterns in metal with positional accuracies of better than 1 μm(ref.14). At 7 MHz in water this corresponds to better than 2^{o} of phase. Such phase accuracies are adequate for radio signal processing and are better than that achieved in most branching network systems.

Simple tests so far have justified the design of a complete acoustic lens system for the radioheliograph. The resultant system should occupy a cubic box less than 50 cm in length and would be equivalent to the five racks of cables which constitute the present branching network. Its loss should be about the same (~25 dB).

Since one has a large degree of freedom in the design of the electrode pattern deposited on the transducer, acoustic processors can be built for most planar arrays. The only major limitation is in the output circuitry. At present no acoustic equivalent to the integrating, self-scanned photodiode array exists. Consequently each output transducer electrode must be individually connected to its own detector and integrating circuit. There might be some hope that in the medical ultrasonics field the concept of applying semiconductor circuitry directly on to the back of the transducer slab might develop to the point where the integrating, selfscanned acoustic detector array is possible.

With such an ideal acoustic detector the acoustic array processor would combine into one low-loss unit not only the functions of delay compensation and image formation but also the processing needed for proper beam formation, detection and integration, and image sampling. Just as in the acousto-optical spectrograph one would then add the flexibility of an on-line digital computer.

7. CONCLUSIONS

The concept of analogue processing clearly has a role to play in imaging with arrays and in radio spectroscopy. The acousto-optical analogue seems preferable for radio spectroscopy and its extensions to interferometry. The purely acoustic analogue offers distinct advantages for imaging array processing. The technology needed to complete the processing in the acoustic device is an integrated acoustic detector with associated read-out circuitry. The result would be a compact, cheap and efficient unit which when combined with a computer could provide real-time images with the full sensitivity of the multi-channel approach.

REFERENCES

No.	Author	Title
1		Collected papers, "The Culgoora Radioheliograph", 1967, Proc. Inst. Radio Electron. Eng. Aust., 9, 277
2	Cutrona, L.J.	"Optical and Electro-Optical Information Processing". (Cambridge, Mass.:MIT Press), Chap. 6, p 83, 1965
3	McLean, D.J. and Wild, J.P.	Aust. J. Phys., 14, 489, 1961
4	McLean, D.J., Lambert, L.B., Arm, M. and Stark H.	Proc. Inst. Radio Electron. Eng. Aust., 9, 375, 1967
5	Cole, T.W.	Opt. Technol., 1, 31, 1968
6	Cole, T.W.	Proc. IEEE, 61, 1321, 1973
7	Cole, T.W.	Astrophys. Lett., 15, 59, 1973
8	Quate, C.F., Wilkinson, C.D.W., and Winslow, D.K.	Proc. IEEE, 53, 1604, 1965
9	Cole, T.W. and Ables, J.G.	Astron. Astrophys., 34, 149, 1974
10	Alphonse, G.A.	Appl. Opt., 14, 201, 1975
11	Cole, T.W.	Opt. Commun., 13, 192, 1975
12	Briggs, B.H. and Holmes, N.	Nature, Phys. Sci., 243, 111, 1973
13	Hart, D.N.	This symposium, 1975
14		"Handbook of Thin Film Technology". Eds. L.I. Maissel and R. Glang, 1970, McGraw Hill, New York

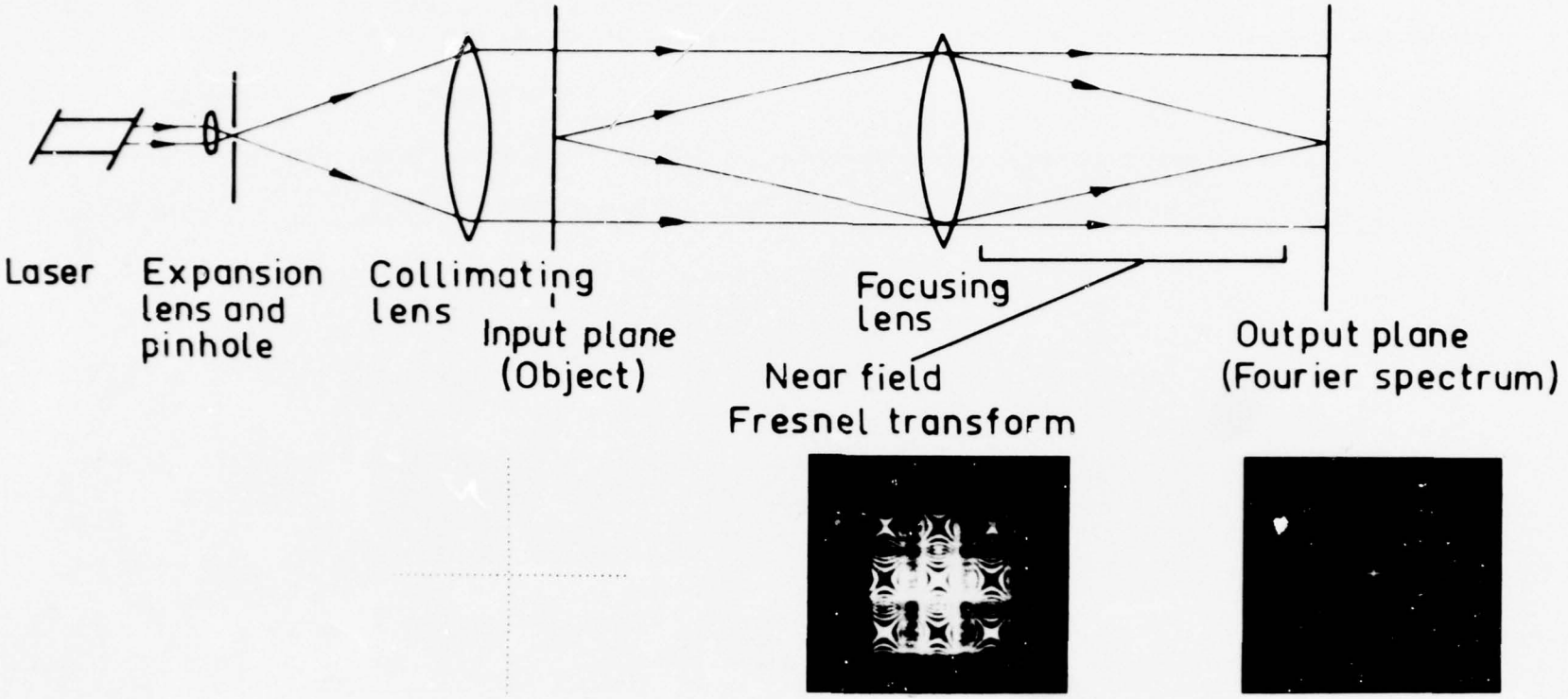


Figure 1. A coherent optical system is shown with an example of an antenna pattern, its Fourier transform, and its Fresnel transform

WRE-MISC-1
Figure 1

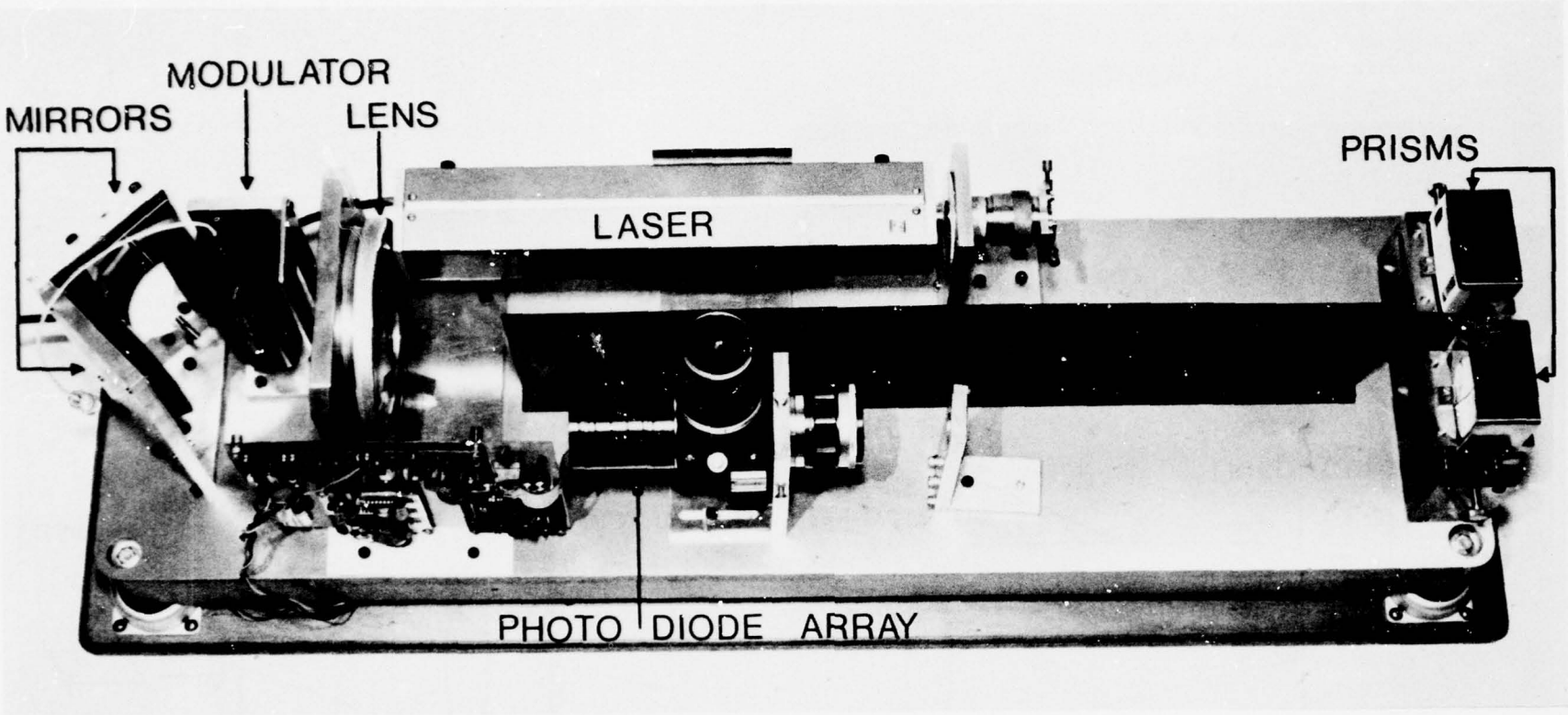


Figure 2. The acousto-optical spectrograph is shown. The base measures 90 x 25 cm and prisms are used to fold the light path

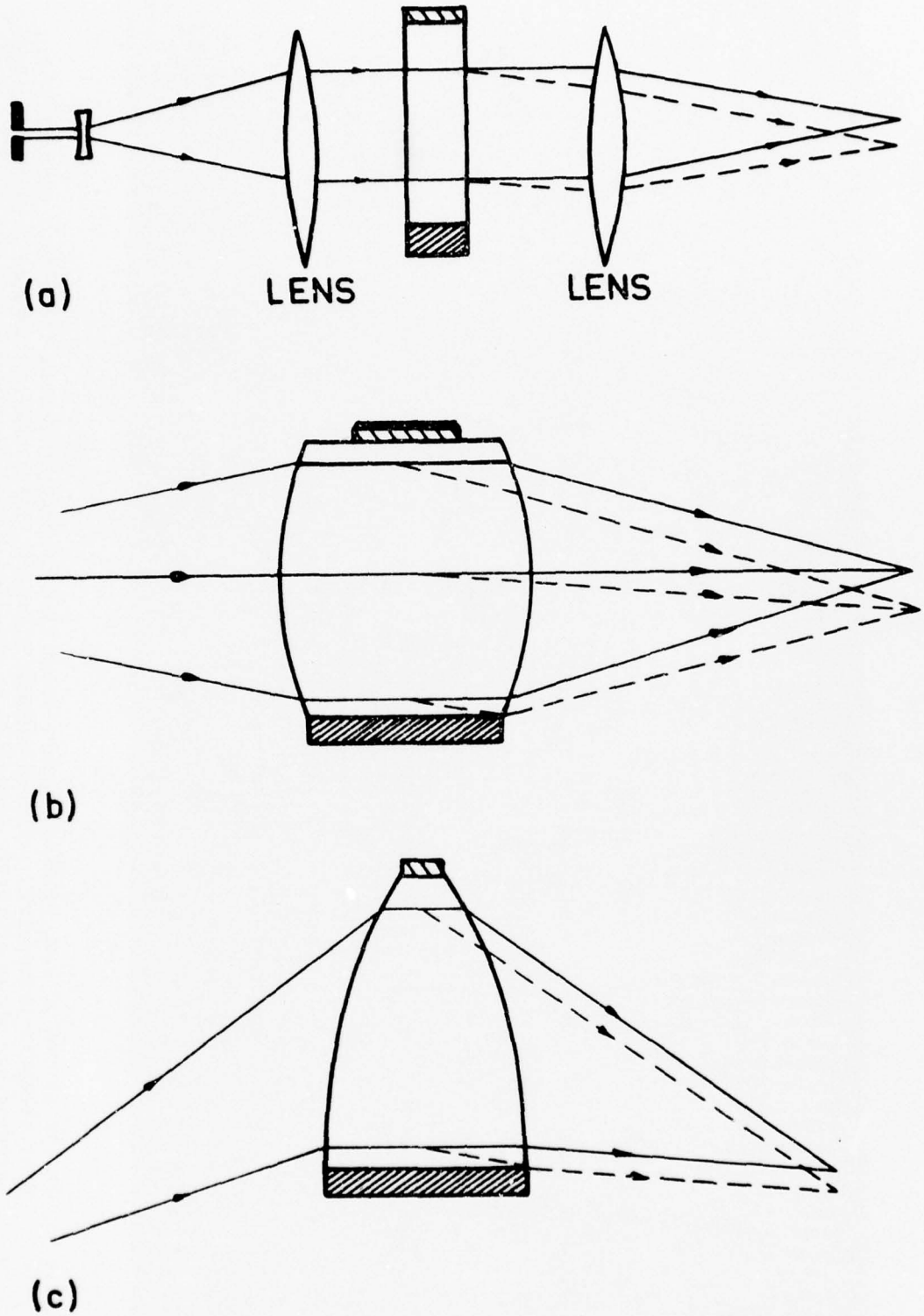


Figure 3. The conventional acousto-optical system is shown in (a) while the combination of two lenses and the modulator is shown in (b). The use of a prism effect to increase the efficiency in illuminating a rectangular sound beam is illustrated in (c)

Figure 4. In the centre of each printed circuit board lead-connection-system there is a circular 7 MHz piezoelectric transducer. The 48-element circular (left) and rectangular (right) electrode patterns are shown partially connected to the circuit board with 25 μm wire and conducting epoxy. The circuit boards are 10 cm x 10 cm

DIGITAL SIGNAL PROCESSING FOR MULTI-ELEMENT ARRAYS

D.F. Crawford, School of Physics, University of Sydney,
New South Wales, 2006

S U M M A R Y

If each signal from a multi-element array is digitised into a single bit response the subsequent processing of these responses can be greatly simplified. The delays required to steer a resultant beam can be achieved with the use of programmable shift registers. The correlation of signals is reduced to simple logical functions and the input of these signals into a computer is easier. By recording both sine and cosine components the phase of each signal need not be known in real time but can in many cases be determined from an analysis of the data. Finally it should be noted that the usual loss in signal-to-noise ratio with single bit digitisation can be reduced by sampling at faster than the Nyquist rate.

Since a significant part of the cost of multi-element arrays is in the electronics required to steer the array and to combine the signals from each element, any method that can simplify the electronics is worth consideration. One such method is to encode the signals into one or more digital bits so that shift registers can replace delay lines, and logical gates can replace analogue multipliers. In many situations the greater simplicity of the one-bit (two-level) digitisation method compared to a multi-bit digitisation method outweighs the loss of signal to noise ratio.

It is not often realised that the degradation in signal to noise ratio of single-bit digitisation compared to the analogue signal to noise ratio can be partly alleviated by sampling the signal at higher than the Nyquist rate. It is shown in the Appendix that the degradation factor of signal to noise ratio for the product of two one-bit signals has an asymptotic value of 1.26 as the rate increases, with most of the improvement reached by three times the Nyquist rate. The degradation factor is much less if one of the signals can be kept in analogue form. For this hybrid case the degradation factor is only 1.15 at twice the Nyquist rate. Cooper's (1970) analysis of a two-bit correlator sampled at the Nyquist rate shows that it has a degradation factor of 1.14 provided the transition voltage is correctly chosen. Although the degradation is not very sensitive to the transition voltage it does make an added complication for multi-bit digitisation if the array has a variable noise level.

The major advantage of the one-bit system over an equivalent analogue system is the simplicity and cheapness of a shift register compared to a switched delay line. By using programmable shift registers (for which the number of shifts are externally controlled) the delay necessitated by the finite bandwidth is easily set for each array element. The setting of phases which accurately steer the array can be postponed for later computation if two signals, one in quadrature, are used and individual cross-products are separately recorded. This has the added advantage that the phases need not be known at the time of observation but could be obtained from the observation of a calibration source. However, real time displays using derived beams would require an on-line computer.

Since the output of a one-bit scheme only gives a measure of the signal to noise ratio, the root mean square noise level must be measured separately. Although this would be done with square-law detectors a better method which checks the operation of the digitisers is to add a low level voltage to each signal, and measure its correlation. If its polarity between two elements is inverted regularly at a rate much faster than a change in signal strength, it can be separated from the true signal by switching the outputs of the multipliers at the same rate.

Another advantage of single-bit digitisation is the simplicity of the signal multipliers which are simple logic gates. Let m_{11}, m_{10}, m_{01}, m_{00} be the counts of the four possible states of two one-bit signals; then the signal power is proportional to $m_{11} + m_{00} - m_{10} - m_{01}$. Provided the signal is small, this summation removes the effects of any biases in the digitisers. The hybrid multiplier is equally simple in that the one-bit signal is used to select either the analogue signal or its inverse as input to an integrator.

As an example of how this one-bit digitisation scheme may be applied to an actual array I have considered its implementation on the North-South arm of the Molonglo Cross (Mills et al, 1963). For this purpose the N-S arm can be thought of as 176 elements, each providing a signal to be multiplied by that from the E-W arm. The centre frequency is 408 MHz and each signal, with a bandwidth of 2.5 MHz, is brought through equal length coaxial cables to the central hut at an intermediate frequency of 11.5 MHz. A digitisation rate of 15 MHz (three times the Nyquist rate) is feasible without using especially high speed integrated circuits, and even if the bandwidth is expanded to a limit of about 4 MHz, sampling rates much higher than the Nyquist rate are still feasible. Clearly a hybrid system is ideal with two analogue signals, one in quadrature coming from the E-W arm to be multiplied by the 176 one-bit signals from each N-S element. The 352 products are integrated and recorded once every three seconds. Since the noise level only changes slowly, the noise calibration outputs need not be recorded as frequently as this. If an on-line computer is used there is considerable advantage in sampling the output more frequently than every three seconds so that some effects of interference can be removed. This computer would also control the shift registers used to steer the telescope in declination. The raw data derived from this system is a spatial Fourier transform (in declination) of the sky brightness distribution recorded as a function of right ascension. The fundamental limitation of the declination covered in one scan is the product of the N-S antenna response of each N-S element and that of the E-W arm. However, the phase gradient within each N-S element is fixed at one particular declination with the result that grating side lobes are formed. Both these factors limit the scan to about 3° in declination. With this width sufficient calibration sources will be seen in a single night's scan to calibrate the phase and absolute gain of each N-S element. Compared to the equivalent analogue system, the hybrid system outline is simpler and has a signal to noise degradation factor of 1.13, that is a 11.5% loss in signal to noise ratio. In this case the added complexity of a two or more bit system is hardly justified.

The author thanks Professor B.Y. Mills for helpful criticism in the preparation of this paper and acknowledges the support of the Australian Research Grants Committee, the Sydney University Research Grants Committee and the Science Foundation for Physics within the University of Sydney.

APPENDIX

For a signal of amplitude x, buried in noise of root mean square amplitude σ, the probability of the output being positive is $\frac{1}{2} + \frac{1}{\sqrt{2\pi}} \frac{x}{\sigma}$. A single-bit digitiser (often called a two-level digitiser or polarity detector) could be used to encode this signal into a string of ones (positive) and zeroes (negative). After a period of time during which x remains constant, n_1 ones and n_0 zeroes are recorded, the best estimate of x is its expected value,

$$< x > = \frac{n_1 - n_0}{n_1 + n_0} \frac{\pi}{2} \sigma ,$$

with variance $\frac{\pi \sigma^2}{2(n_0 + n_1)}$. When the sampling rate is not at the Nyquist frequency (which is twice the bandwidth), allowance must be made for correlations between samples. The analysis for the product of two signals has been done by Hagen (1972) using a treatment similar to Burns and Yao (1969). If m is the relative sampling rate so that it is the Nyquist rate when m = 1, the signal to noise degradation ratio for three cases of interest are

(1) Product of two analogue signals

$$\beta_1 = \frac{1}{m}(1 + 2 \sum_{k=1}^{\infty} \rho_k^2),$$

(2) Product of analogue by 2 level digital (hybrid)

$$\beta_2 = \frac{\pi}{2m}(1 + \frac{4}{\pi} \sum_{k=1}^{\infty} \rho_k \text{ arc sin } (\rho_k)),$$

(3) Product of 2 level digital by 2 level digital

$$\beta_3 = \frac{\pi}{2m}(1 + \frac{8}{\pi^2} \sum_{k=1}^{\infty} (\text{arcsin } (\rho_k))^2),$$

where ρ_k is the correlation coefficient between samples k intervals apart, and it is assumed that the signal changes very slowly compared to the sampling rate. The major difference between β_1 and β_2 (or β_3) is the incorporation of the Van Vleck relation (1966) which relates the correlation ratio of a two-level output to the correlation ratio of the original signal. The correlation coefficients ρ_k are obtained by taking the Fourier cosine transform of the noise power spectrum. For a rectangular bandpass filter the result is $\rho_k = \text{sinc}(\frac{k}{m})$. Table 1 lists values of β_1, β_2 and β_3 as a function of m for this case of a rectangular noise spectrum. A few calculations with other spectra suggest that the degradation ratios are not very sensitive to the spectral shape provided the total power remains the same.

TABLE 1

m RATE	β_1 AXA	β_2 AX2	β_3 2X2
0.5	1.41	1.77	2.22
0.6	1.34	1.66	2.06
0.7	1.26	1.55	1.92
0.8	1.18	1.45	1.80
0.9	1.10	1.35	1.68
1.0*	1.00	1.25	1.57
1.2	1.00	1.21	1.49
1.4	1.00	1.19	1.43
1.6	1.00	1.17	1.39
1.8	1.00	1.16	1.36
2.0	1.00	1.15	1.34
2.5	1.00	1.14	1.31
3.0	1.00	1.13	1.29
3.5	1.00	1.12	1.28
4.0	1.00	1.12	1.28
5.0	1.00	1.12	1.27
6.0	1.00	1.12	1.26
8.0	1.00	1.12	1.26
10.0	1.00	1.11	1.26

* Nyquist frequency

REFERENCES

No.	Author	Title
1	Burns, W.R., and Yao, K.S.	"Clipping loss in the one-bit autocorrelation spectral line receiver". Radio Science 4, 431-436 (1969)
2	Cooper, B.F.C.	"Correlators with Two-bit Quantization". Aust. J. Phys. 23, 521-527 (1970)
3	Hagen, Jon B.	"A Hybrid Autocorrelator and its application to High Altitude incoherent scatter". National Astronomy and Ionosphere Center, Ithaca, N.Y. Report 18, 1972
4	Mills, B.Y., Aitchison, R.E., Little, A.G. and McAdam, W.B.	Proc. Inst. Radio Engineers (Aust.) 24, 156 (1963)
5	Van Vleck, J.H. and Middleton, D.	"The Spectrum of Clipped Noise". Proc. I.E.E.E. 54, 2-19 (1966)

AN ULTRASONIC IMAGING SYSTEM FOR USE WITH A LARGE AERIAL ARRAY

D.N. Hart, Department of Physics,
University of Adelaide

SUMMARY

An ultrasonic imaging system has been constructed for use with the large aerial array at the University of Adelaide field station at Buckland Park. This system allows direct observation of the modulus of the angular spectrum of radio energy impinging on the aerial array at either of two frequencies (2 MHz and 6 MHz). The analysis is performed in real time.

A description of the equipment is presented, along with some applications in ionospheric physics and low frequency radioastronomy.

1. INTRODUCTION

The imaging system described in this paper has been constructed at the University of Adelaide field station at Buckland Park (approximately 40 km north of the city) for the purpose of ionospheric and radioastronomical research. It is designed to allow observation in real time of the angle of arrival of radio energy at 2 or 6 MHz incident upon a large antenna array. The system is limited to pulsed operation for reasons that are discussed in Section 3.2.

2. THE ANTENNA ARRAY

The main receiving array consists of 178 crossed dipole antennas (figure 1) which can either be connected to receive circularly polarized radiation, or operated separately to form two independent 89 element linearly polarized arrays. The electrical length of the cable from each antenna is adjusted so that the relative phase between signals produced in different aerials is preserved in the central laboratory at both operating frequencies of 1.98 and 5.995 MHz.

A pulse transmitter located at the edge of the array provides linearly or circularly polarized transmissions at 1.98 MHz (which will be referred to as 2 MHz) with a maximum peak power of about 100 kW. Transmissions at 5.995 MHz (6 MHz) are, at the present time, restricted to linear polarization with a peak power of 4 kW. A more detailed description of the array is given by Briggs et al (1969).

3. THE IMAGING SYSTEM

3.1 Brief outline

A block diagram of the imaging system is shown in figure 2. Its operation is as follows.

A radio diffraction pattern (of ionospheric or radioastronomical origin) formed over the ground is sampled by the 89 aerials of the array. These signals are amplified, converted to an IF frequency of 455 kHz, and fed to an array of piezoelectric transducers at one end of a large stainless steel water-filled tank. The transducers are arranged in the same pattern as the aerials of the array so that an acoustic wavefront similar in amplitude and phase to the original radio diffraction pattern is propagated through the water. The transducers are, however, arranged on the surface of a sphere so that the wavefront produced is curved and comes to a focus at a distance equal to the radius of curvature of the transducer array. A second set of transducers is arranged as a receiving array in the focal plane of the transmitting array (at a distance of 1.65 m) and samples the acoustic image produced. The signals from this set of transducers are fed to amplifying and detecting circuits which drive a light emitting diode display that gives an optical presentation of the focused acoustic image. This then represents the modulus of the angular spectrum (or intensity vs angle of arrival) of the radio energy incident upon the aerials of the main array.

3.2 Discussion

It is critical for the successful operation of the instrument that the relative phases of the 89 sampled signals must be kept constant throughout the amplifying and mixing stages and up until they are fed into the water tank. This is achieved by operating all of the superheterodyne receivers at constant gain so that the phase shift through them is stable (the phase shift is gain dependent). Also, all of the mixers are driven by a common local oscillator ensuring correct phase information in the IF signal.

Constant gain operation of the receivers seriously degrades the dynamic range of the system so, to remedy this, a set of voltage controlled attenuators prefaces the set of receivers. Each of these is driven by a voltage derived from the average level of the corresponding receiver output in such a way as to maintain the output at an approximately constant level. The time constant is variable but is usually set at about 100 s. The attenuators have a range of -1 to -35 dB. This arrangement, along with a facility for varying the transmitted power, provides sufficient dynamic range to cover most input signal amplitudes. It might, at first sight, appear that this method would seriously degrade the signal to noise ratio of the system. However, as most of the noise is externally generated, this is not so.

Water was chosen as the acoustic medium for the imaging system because

(1) it provides a reasonable impedance match for the transducers (good energy transfer from transducer to water,

(2) the wavelength of 455 kHz ultrasound in it (3.2 mm) is a convenient length, and

(3) it is cheap.

As noted in Section 1, the imaging system is restricted to pulsed operation. This restriction is due to the fact that the stainless steel surfaces of the tank are excellent reflectors of the acoustic waves and it is necessary to ensure that these unwanted reflections do not interfere with the acoustic image. This is achieved by pulsing the input signal to the tank. The shortest possible time interval between the arrival of the wanted image and the first reflected signal is slightly more than 100 μs. In practice, the most significant wall reflections arrive about 160 to 200 μs later. The time for all reflections to dissipate after a pulse is of the order of 5 ms. Thus, provided the input pulse is shorter than about 100 μs and the repetition frequency is not higher than 200 Hz, no interference from reflections will be encountered. The pulsed input to the tank is obtained by pulsing the common local oscillator. This, in effect, switches the receivers on for the required interval and suppresses their output at all other times. In ionospheric applications the pulse is adjustable both in width and delay so that, for example, a particular ionospheric echo may be selected. In the case of continuous inputs (e.g., radioastronomical observations) the pulsed local oscillator serves to "chop" the input signal so that it may be processed by the system.

When the receiving array is operating at 2 MHz the field of view of the imaging system is $\pm 28^{o}$ from the zenith with a resolving power of approximately $\pm 4.5^{o}$ to the half power points. At 6 MHz, the resolving power is increased by a factor of 3 but the field of view is correspondingly decreased.

4. APPLICATIONS

The imaging system can provide unambiguous information on the direction of arrival of radio energy under all conditions. In this respect it is superior to most direction finding systems which, in general, fail when energy is incident from several directions simultaneously. Ionospheric irregularities which produce this type of reflection are not well understood. For example, structures which produce such complex reflections are spread F irregularities, sporadic E formulations, and D-region irregularities. The imaging system described here should prove a valuable tool in leading to an understanding of these phenomena.

Radioastronomical observations below 10 MHz are rare and directional information at frequencies less than 10 MHz is non-existent. Measurements at these low frequencies are generally only possible during years of sunspot minimum when the ionospheric critical penetration frequency falls to a sufficiently low value. We are currently in such a period and it is planned to use the imaging system to carry out a sky survey at 6 MHz. The resolution of the survey would be approximately 1^{o}.

REFERENCES

No.	Author	Title
1	Briggs, B.H. et. al.	Buckland Park Aerial Array, Nature 223, 1321, (1969)
2	Homes, N.E.	An Ultrasonic Image-forming system for use in conjunction with a large Aerial Array. University of Adelaide, Department of Physics, Ph.D. Thesis, (1974)

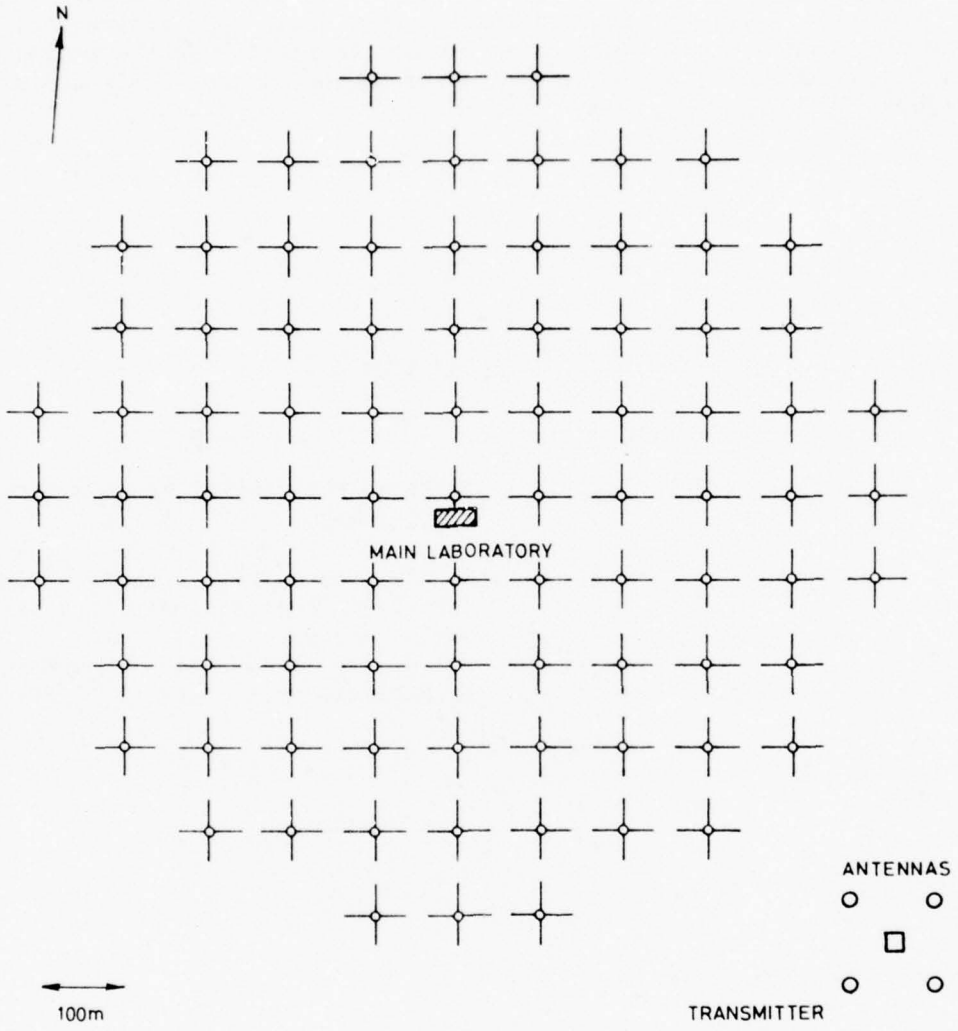


Figure 1. Diagram showing the layout of the receiving array. Each cross represents a pair of dipoles. The transmitting array is positioned 700 m to the south-east of the array centre.

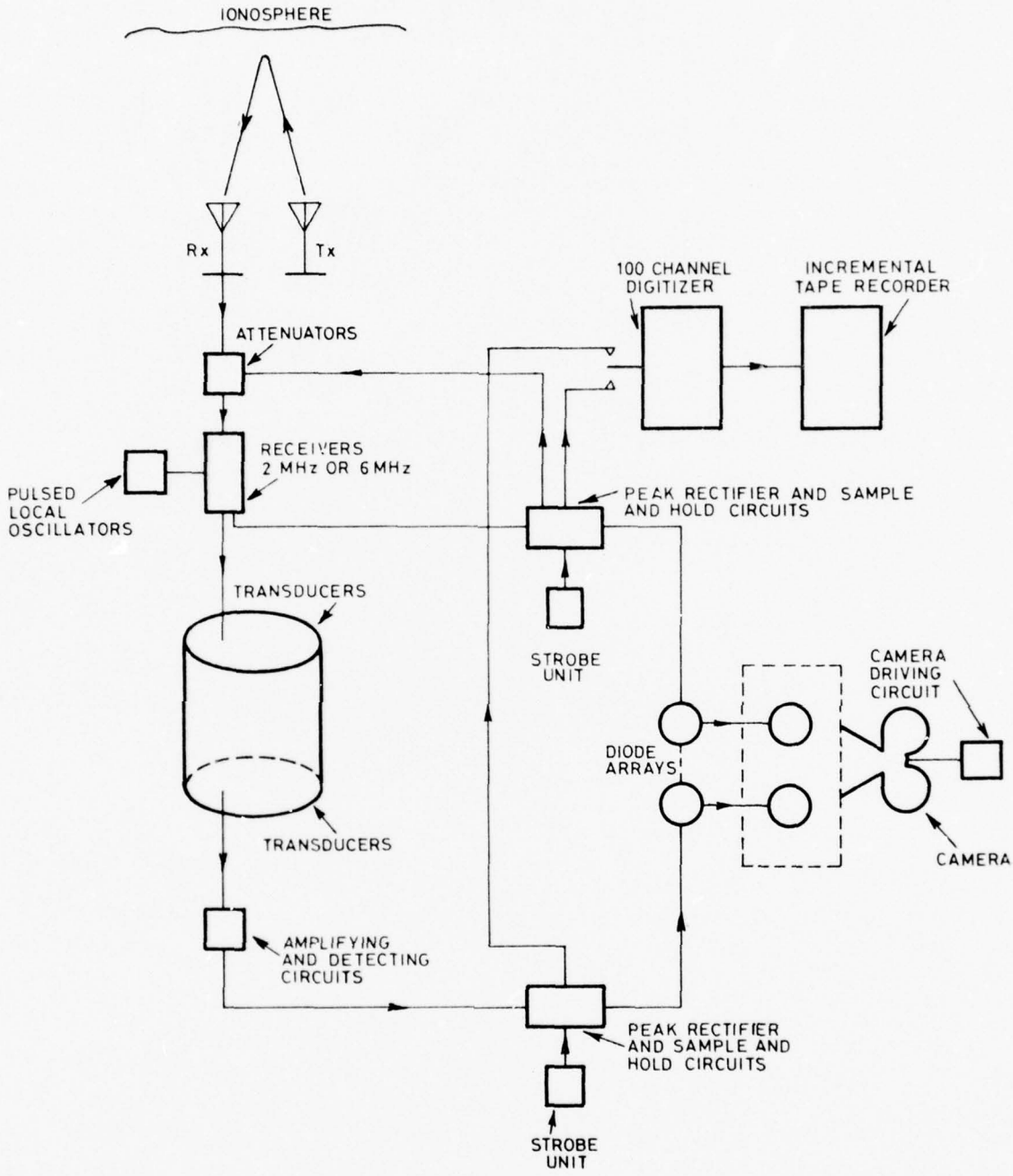


Figure 2. Detailed block diagram of image-forming system

SIGNAL PROCESSING IN SEISMOLOGY

Gordon Newstead

Emeritus Professor, Australian National University

1. INTRODUCTION

This paper presents a review of processing methods as currently used in array seismology. It is intended for an audience interested in the general principles of array processing. The material presented draws heavily on a forthcoming book by Dr Muirhead and myself. As will be evident from the papers presented at this symposium, whilst there is considerable similarity at a fundamental level between the methods of array processing used in the various fields, there are also considerable differences. My original interest in seismic processing arose partly out of an acquaintance with radio astronomy and the realization that the correlation techniques developed there could be applied to seismology. In the event, very few of these techniques could be taken over directly, the differences being due to the following factors:

(1) the characteristics of the signal (transient nature, long wavelength, low frequency),

(2) the characteristics of the noise,

(3) technological difficulties.

The transient nature of the signal (for example, the p-wave lasts only a few seconds) means that long integration times cannot be used as a method of signal/noise enhancement. The long wavelength (the order of 10 km) means that the difficulty and cost of constructing an array of even medium aperture is very great. The low frequency creates two sorts of difficulties. Firstly, when these studies were beginning the technological difficulties of constructing stable low-noise d.c. amplifiers were considerable and seismology, as did radio astronomy in its early days, went through a period where a great deal of its effort was devoted to the design of the front end of the system. Fortunately with the development of transistor technology this problem has receded, at least for short period seismology which is dealt with here. Long wavelengths and low frequencies also implies low propagation speeds ($\sim$ 10 km/s) and this inhibited the development of on-line analogue processing. The full development of seismic processing had to await the time when recorded versions of the signals were available in a form suitable for directly feeding into a computer. This too brought its problems, as most of the records are of seismic noise which is of little intrinsic interest (although this is not known until after the records have been processed), so that the system tends to be flooded with irrelevant data.

A difficulty that arises in applying conventional noise theory to seismology is that this theory is mainly developed for ergodic random sources, whereas a large component of seismic noise possesses spatial coherence. This of course makes it easier to discriminate against the noise, but means that some of the algorithms developed by seismologists for dealing with their signals may not be applicable directly in other fields.

A further bugbear in the processing of seismic arrays is the occurrence of noise of non-seismic origin. Some of these extraneous noises which get into the system possess coherence over the whole array and thus can masquerade as signals. For example, lightning which can affect all the seismometers in an array simultaneously will be interpreted as a seismic signal arriving from the direction of the downward vertical unless special methods are adopted to recognise it.

The continuing solution of the technological problems discussed above are at least as important to seismic array processing as the development of processing algorithms of the type dealt with in this paper.

In the paper we discuss some of the processing methods more commonly used for arrays. Non-linear methods are only dealt with lightly as they will be covered by other authors at this symposium and well known techniques such as deconvolution, frequency filtering, etc., are only treated incidentally.

2. THE GENERAL SEISMIC ARRAY

A seismic array consists of a number of sensors (seismometers) spread out over the surface of the ground, generally in a geometrical pattern, and with their outputs processed in such a way as to take account of the coherence of the wanted signals. Systems in which the unprocessed output of a number of telemetered seismometers are presented synchronously to a seismic interpreter are not here considered as arrays, although it must be pointed out that a skilled interpreter carries out intuitively a great deal of 'eyeball' processing and the results which can be achieved by this method are often surprising. Nevertheless, the objective of array processing is to automate and refine the process of interpretation and to present the output in a form where it can be interpreted with less skill or whereby information which is otherwise not apparent can be obtained. A great deal of the effort in seismic processing has gone into the presentation of reconsituted seismograms in which the signals can be studied free from interfering noises. Some of the processing, however, results in direct computer output of the desired parameters of the signal.

The long wavelength of seismic signals means that the array is spread over a large piece of land. The U.K.A.E.A. type array at Tennant Creek is approximately in the form of a cross with arms 25 km long. LASA (the 'Large Aperture Seismic Array', Green et al 1965) in Eastern Montana, U.S.A., has an aperture of 200 km.

Experiments have been carried out with spatial arrays, the seismometers being located in deep boreholes. However, because a large component of the unwanted noise consists of surface (Rayleigh) waves, very deep holes are required to effect much improvement and it turns out to be more economical to build a larger array on the surface of the ground.

An M sensor array is shown in figure 1. It consists of M seismometers spread out over the surface of the ground, generally in a geometrical pattern as well as amplifiers, telemetry equipment, transmission links, recorders and a signal processor. The signal processor may range between simple bandpass and band splitting filters to on-line computers both for editing what is to be recorded and for automatic event detection. As the ultimate sensitivity of an event detector depends, amongst other things, on the amount of time spent in the processing, the amount of preprocessing which is carried out locally or on-line is always a compromise between the amount of information it is desired to extract from the signal and the cost and nuisance of having to record all the time when it is known that most of the record will be of seismic noise.

Because seismometers and their associated equipment are expensive, seismic arrays have a small number of elements: the U.K.A.E.A. twenty, and LASA, the largest in the world, now has 326 (this number was originally 525 but because the extra seismometers were not fully effective they were removed for another purpose). Perhaps it would put this size aspect into perspective by describing a U.K.A.E.A. type array as the seismic analogue of a Mills cross with an aperture of approximately 2λ. This is a very small aperture by the standard of arrays used in other fields of study. It is for this reason that seismologists have had to turn their attention to the most sophisticated processing methods in order to extract the maximum amount of information.

Given digital processing, the exact array geometry is not so important except in determining the magnitude of the computing task. In the past when delays were inserted by tape loops or when, because of limitations on computer capacity, a certain amount of preprocessing was done by analogue methods, it helped if the seismometers were arranged in geometrical patterns, for example, straight lines. The usual discussions, as dealt with in antenna theory about how to dispose the elements so as to achieve given lobe and side-lobe patterns, apply. However the question: "What is the optimum distribution of M seismometers?" does not have an answer as the distribution depends on the purposes for which the array will be used.

3. SIGNAL PROCESSING

We shall assume off-line processing on a recorded version of the output of the M seismometers. Processing in real time does not involve any great differences except that even then there will be, inevitably, time delays between the receipt of the signal and the availability of the output.

A general theory of linear seismic processing has been given (Green et al 1966). This will now be outlined and special cases discussed.

The seismic signal is specified by an angular frequency ω and a vector wavenumber $\vec{k}$. Signals of various types can thus be represented as lying in various parts of ω-$\vec{k}$ space. The properties of the array are represented by a function $G(\omega,\vec{k})$ having particular values at points in this space - $G(\omega,\vec{k})$ will be recognised as the directivity function of antenna theory. By shaping $G(\omega,\vec{k})$ (that is, by choosing the array geometry and the processing method) so that it is large in those parts of ω-$\vec{k}$ space where the desired signal is located and small in the regions where the noise resides, an array can be 'tuned' to have an optimum response to the desired signal - for example, a lobe can be pointed towards the signal or a null towards the noise. Further, for wide-band systems, this can be done frequency by frequency. It can be shown (Christiansen and Hübgom, 1969) that $G(\omega,\vec{k})$ is given simply as the spatial Fourier transform of the 'illumination' across the array.

Let $\vec{U}(\omega,\vec{r})$ be the movement of the ground at the point $\vec{r}$. Let the ground movement be weighted by a factor $W(\vec{r})$ and the signal delayed by $\tau(\vec{r})$ then the $G(\omega,\vec{k})$ which results is given by

$$G(\omega,\vec{k}) = \int_{\substack{\text{whole}\\ \text{array}}} W \exp\left[j(\omega\tau + \vec{k}.\vec{r})\right]\vec{U}\, d\vec{r}. \qquad (1)$$

In this equation G is normalised with respect to the output of a single seismometer, rather than to unity as is usual in antenna theory. We make the simplifying assumption that, apart from a phase (time) difference, $\vec{U}(\omega,\vec{r})$ is the same at all the seismometers which are localised at points on the plane. Thus we can rewrite equation (1) as

$$G(\omega,\vec{k}) = \sum_{\substack{\text{all}\\ \text{seismometers}}} W(\vec{r})\exp\left[j(\omega\tau + \vec{k}.\vec{r})\right] \qquad (2)$$

In practice we are interested in $|G|^2$; that is, power rather than amplitude and this is simply obtained from equation (2). Applying the process described in equation (2) is analogous to beamforming and we shall now apply this to various situations.

4. CLUSTER ARRAYS

A cluster was the first type of array to be developed and consists of a number of seismometers whose outputs are simply summed. This corresponds to $W(\vec{r}) = 1$, $\tau = 0$ in equation (2), from which it can be seen that $|G|^2$ has a maximum value of M^2 when $\vec{k} = 0$; that is, when the beam points vertically downwards. For random noise signals which arrive equally from all directions at a linear array with elements equally spaced half a wavelength apart, it can be shown from equation (2) that the average value of $|G|^2$ is M. Thus such a cluster will give a power gain of M in signal-to-noise for signals which arrive from the direction of the downward vertical and approximately this gain for signals which arrive 'near' this direction: in this context 'near' means angles of arrival for which the apparent wavelength over the surface of the ground is large compared with the dimensions of the cluster. It should be pointed out for those not familiar with seismic wave propagation in the earth that because the velocity of seismic wave propagation increases with depth in the earth the first arrivals from distant events come from a direction near the downward vertical.

5. DELAY AND SUM

The obvious improvement to make to a cluster is to introduce delays into the signals before summing so as to give freedom to point the main lobe in a direction other than vertically downwards. Turning again to equation (2) we put in this case $W(\vec{r}) = 1$, $\tau = -\vec{k}_o.\vec{r}/\omega$ and obtain

$$G(\omega,\vec{k}) = \sum_{\substack{\text{whole}\\ \text{array}}} \exp[\, j(\vec{k}-\vec{k}_o).\vec{r}\,] \,. \qquad (3)$$

This is the same expression as for a cluster except that $\vec{k} - \vec{k}_o$ replaces $\vec{k}$ and the main lobe can now be pointed, frequency by frequency, in any direction $\vec{k}_o$. A further refinement can be achieved by weighting the elements in a delay and sum array. This introduces the possibilities of the usual compromise between suppressing sidelobes at the expense of widening the main beam and, if the weights are also made functions of ω, of frequency filtering.

An example of what can be done by delay and sum processes is shown in figure 2. The figure shows various outputs from an event recorded on a U.K.A.E.A. type array at Tennant Creek and processed digitally by Dr Muirhead at the ANU. The record also includes some other processing which will be referred to later. The trace on the first line is the output of a typical seismometer in the array. The trace on the fourth line shows the phased and summed outputs of all the 20 seismometers in the array. The existence of a seismic signal arriving from the direction to which the array is tuned is evident on the fourth trace whereas this is not apparent on the first trace which looks like a typical record of random seismic noise.

6. MAXIMUM LIKELIHOOD FILTERING

In order to explain the objectives of this method of filtering, we start with a digression about how the subject of seismic array processing developed. When work commenced in these areas, digital computers and their ancillary equipment were not as developed or available as they are today and even now the amount of information to be processed presents a formidable task. The inflexibility of analogue devices once they had been built, as well as the need to restrict the amount of data processed by those who had access to digital equipment, led to the development of two schools of thought conditioned both by seismological considerations and the purposes for which the information was required. The big drive and large amount of money available for the development and construction of seismic arrays originated from the desire to monitor underground nuclear explosions with a view to policing an underground nuclear test ban treaty. In nuclear surveillance one can distinguish two distinct problems; firstly, the detection of suspicious events and secondly their positive identification either as earthquakes or explosions. Briefly one can say that given good, sufficiently noise-free records there are seismic criteria which enable this discrimination to be carried out, but however much the equipment and processing methods are refined there always will be events which can be detected but which cannot be identified positively. For historical reasons, this led one group (mainly in the U.S.) to concentrate on the production of more noise-free records by linear processing methods which do not alter the 'signature' of the original seismogram. The other group (mainly the U.K.A.E.A.) put more effort into processes that increase the detectability of signals in noise. The first of these philosophies led to the development and application at LASA of the maximum likelihood method of processing and the second to the correlation processes used in the U.K.A.E.A. array. Of course, both groups now have the capacity to apply either method as they wish.

Maximum likelihood filtering is a special case of filter, weight, delay and sum in which restraints are placed on the filter coefficients by the requirement that the signals be undistorted. It can be shown (Green et al 1966) that the most general case of filter, weight, delay and sum is achieved if running samples of each seismometer output are taken over a time interval, T. These samples are then weighted, delayed, summed and averaged over T. By adjustment of the weights and the sampling time this process includes straight sum, delay and sum, and filter, weight, delay and sum: it is in fact the most general linear processing which can be carried out.

Suppose that the noise is a multidimensional Gaussian process and that the signal, except for propagation delay, is the same at every output. The output of the i-th seismometer is sampled at uniform discrete time intervals t and it is assumed that the appropriate time delays to bring the signals into phase have been inserted. The output of the i-th seismometer at time $r\tau$ may be expressed as

$$\ell_i(r) = s(r) + n_i(r) \tag{4}$$

where $s(r)$ is the signal and $n_i(r)$ the noise. A set of L filter coefficients for every seismometer can then be defined for the array processor, that is, there are LM filter coefficients. The output $y(r)$ of the processor at time $r\tau$ is given by

$$y(r) = \sum_{i=1}^{M} \sum_{k=0}^{L-1} \ell_i(r+k)\omega_i(k). \tag{5}$$

In order to pass the signal which is common to all seismometers without distortion whilst at the same time suppressing the noise as much as possible, Levin (1964) introduced the constraints

$$\sum_{i=1}^{M} W_i(0) = 1$$

$$\sum_{i=1}^{M} W_i(k) = 0, \quad \text{for } k \neq 0.$$

That this does not change the waveform of the signal, may be shown by interchanging the order of the summation in equation (5) which, after some manipulation, gives:

$$y(r) = s(r) + \sum_{k=0}^{M} m_i(r-k)\, W_i(k).$$

The filter coefficients are normally determined on a section of the array record just prior* to the signal onset so as to minimise the average value of $[y(r) - s(r)]^2$. In the language of statistical theory, this least square criterion yields a maximum likelihood estimate of the signal and, given the constraints, no linear process can be more efficient. A matrix method for calculating the coefficients and carrying out the process has been given by Claerbout (1968). If the spacing between the seismometers is sufficient to render the noise incoherent between them, maximum likelihood processing will degenerate into simple delay and sum and yield a gain of M. An interesting question is whether a signal/noise improvement of greater than M can be obtained with this method if the seismometers are moved closer together so that the noise becomes partially coherent. Clearly in cases where part of the noise is highly organised this will be so and cases of improvement greater than M have been reported in the literature (Bachus et al 1964). Based on experiments with data from LASA, in which delay and sum processing over large apertures were compared with maximum likelihood filtering for a smaller sub-array, it appears that practically the same signal/noise improvement can be obtained by either method. Thus the main advantage of maximum likelihood processing over delay and sum is in suppressing coherent energy from interfering events and for improving a poor array site when logistic or other reasons prevent a better site from being used.

7. CORRELATION PROCESSING

If we give up the requirement of the preservation of the waveform of the signal and turn to the detection of signals, the whole field of non-linear processing is available. Here we deal with correlation processing. Other methods of non-linear processing are discussed by other authors at this symposium.

Firstly we consider the process of correlating two signals and how this is applied to the U.K.A.E.A. array. Turning to equation (4) for i = 1 and 2 we define $\phi_{12}(t,\tau,T)$ as

$$\phi_{12}(t,\tau,T) = \frac{1}{T}\int_{t-T}^{t} i_1(t')\, i_2(t'-\tau) W(t')\, dt'. \qquad (6)$$

---

* Note that this makes the process an adaptive one but there is evidence that this aspect does not result in any improvement.

The time T is called the window of correlation and $W(t')$ is a weighting factor. If $W = 1$ the window is called square; we shall confine ourselves here to this case, although other windows are used for particular purposes. For example an exponential window is more useful in determining onset times because it gives most weight to the most recent information. It can be shown (Birthill and Whiteway, 1965) that if B is the bandwidth of the signal used in this correlation process and S/N the original signal/noise ratio then the signal/noise ratio $(S/N)'$, produced by the process defined by equation (6) will be

$$\left(\frac{S}{N}\right)' = \left(\frac{S}{N}\right)^2 (2BT)^{\frac{1}{2}} / \left(1+2\left(\frac{S}{N}\right)^2\right)^{\frac{1}{2}}. \qquad (7)$$

Because of the low values of 2BT ($\simeq 4$) which can be used, an improvement in signal/noise only results from this process if $S/N \simeq 1$ and we then have

$$\left(\frac{S}{N}\right)' \simeq (BT)^{\frac{1}{2}} \left(\frac{S}{N}\right).$$

The correlator signal/noise may be further improved by dividing the seismometers into two almost equal groups, phasing and summing each group and cross correlating the sums. If there are M/2 seismometers in each group, this process increases the signal/noise ratio by a further M/2. A further improvement of about 3 db is theoretically possible for small signal/noise ratios by cross correlating every seismometer pair in the array, but in practice the small values of signal/noise for which this improvement would be available are not usable. It is pointed out that in seismic array processing a 3 db improvement is well worth having; to obtain this amount of improvement by increasing the size of the array would involve roughly doubling the size (and cost) of the array.

The U.K.A.E.A. array was designed for processing by off-line analogue computer as above. It consists of two lines of seismometers roughly in the form of a cross although topography often causes deviation from this geometry. A description of this array has been given (U.K.A.E.A., 1965).

As an example of what can be achieved by this method we return to figure 2 where the fifth trace shows the correlator output. The great improvement over a single seismometer is observed; although this record does not demonstrate much advantage over delay and sum processing, one can feel more confident about the arrival time of coherent energy.

8. SIZE OF COMPUTING PROBLEM

To conclude this survey of array processing, we consider two aspects of the amount of data that has to be handled in seismic array processing, recording and on-line beamforming.

With digital recording, a standard half-inch, 2 400-foot magnetic tape sampled at the rate of 20 Hz for 20 channels (U.K.A.E.A. type array) with a packing of 800 bits/inch would fill in less than 6 hours. For LASA the equivalent time would be about 10 minutes. In fact, the U.K.A.E.A. arrays use analogue recording on a one inch, 7 200-foot tape which lasts 10 days. For LASA which uses digital recording there is not only the problem of supplying these tapes but also of processing them. This leads to consideration of the process of continual surveillance by on-line beamforming. If the whole of the third zone were to be searched by LASA the number of beams required would be greater than 40 000, requiring prodigious on-line computing capacity. Because of the large number of beams required, even for a medium aperture array, the geometry of the array is usually partly determined so as to reduce the

processing required. Dividing the whole array into sub-arrays (such as the two straight lines in the U.K.A.E.A. array or the 2 sub-arrays in LASA) reduces the processing problem by allowing the use of some analogue processing (filtering, summing), but it must be recognised that these procedures required to achieve economy in processing prevent the attainment of the ultimate theoretical capability of the array.

9. CONCLUSION

This paper has discussed the philosophy of the methods of seismic array processing which have been developed with particular reference to linear and cross correlation processes. Whilst in the future processes will be developed which will be more relevant for particular problems or more efficient in their use of computing time, the author does not expect to see further advances comparable with those which resulted from the introduction of the digital computer into this area.

REFERENCES

No.	Author	Title
1	Bachus, M., Burg, J., Baldwin, D. and Bryan, E	'Wide band extraction of mantle p waves from ambient noise' Geophysics 29 672 - 692, 1964
2	Birthill, J.W., and Whiteway, F.E.	'The application of phased seismic arrays to the analysis of seismic body waves' Phil. Trans. Roy. Soc., 258 421 - 493, 1965
3	Claerbout, J.F.	'A Summary by illustrations of least squares filters with constraint' I.E.E.E. Trans in Information Theory Vol. IT-14 No. 2, p 269-272, 1968
4	Green, P.E. Jr., Frosch, R.A. and Romney, C.F.	'Principles of an experimental large aperture seismic array (LASA)' Proc. I.E.E.E. 1821 - 1833, 1965
5	Green, P.E., Kelly E.J., and Levin, M.J.	'A comparison of seismic array processing methods' Geophysics J.R.A.S. 11 67 - 84, 1966
6	Levin, M.J.	'Report on seismic discrimination' M.I.T. Lincoln lab. semi-annual tech. summary, (December, 1964)
7	United Kingdom Atomic Energy Authority Special Report	'The detection and recognition of underground nuclear explosions' H.M.S.O. London, 1965
8	Christiansen, W.N., and Höbgom, J.A.	'Radiotelescopes' Cambridge Monograph on Physics, Cambridge University Press, 1969.

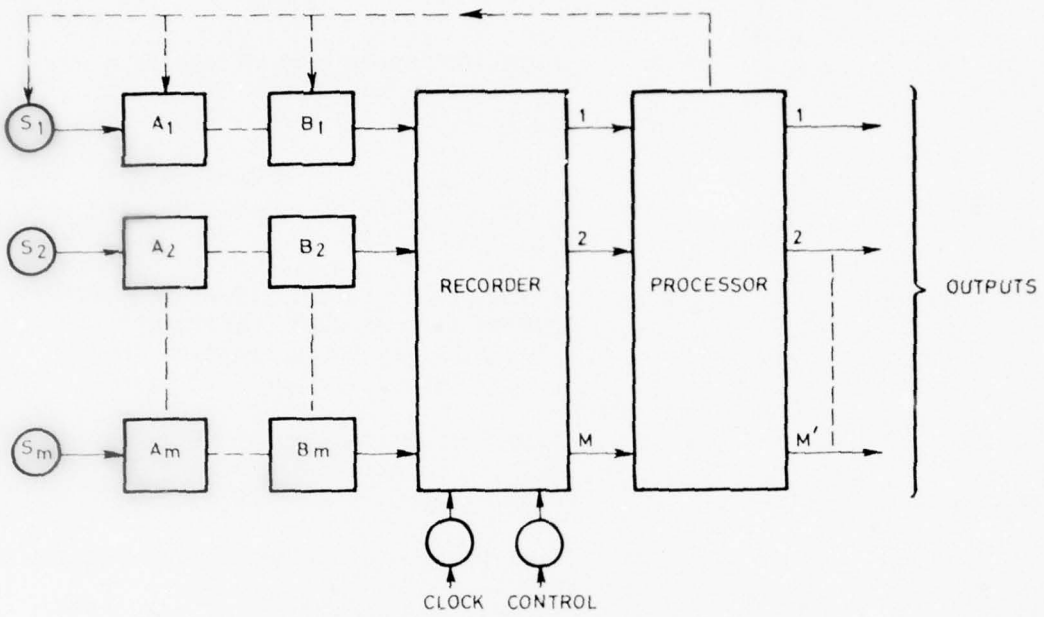


Figure 1. General arrangement of an array

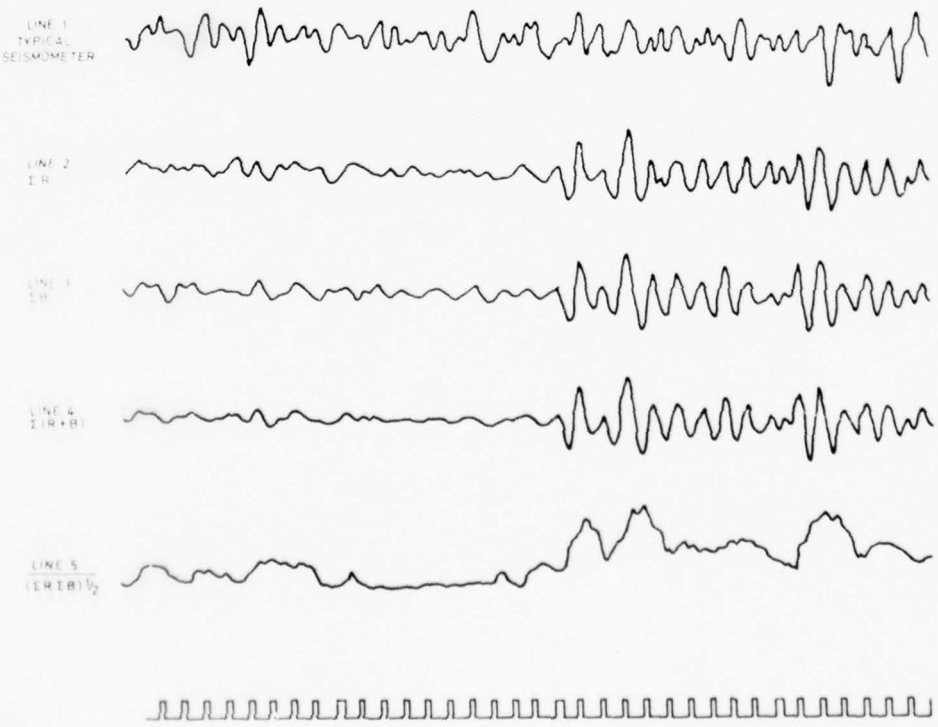


Figure 2. Outputs using different types of processing

NTH ROOT PROCESSING OF SEISMIC SIGNALS

K.J. Muirhead & Ram Datt, Research School of Earth Sciences
Australian National University, Canberra, A.C.T.

S U M M A R Y

A process of taking the Nth root of individual signals from a seismic array before combining them into array beams has been found to have distinct advantages over linear processing in both the detection of signals and the determination of their direction of arrival.

Although the improvement is largely due to this process being better able to cope with non-gaussian noise and signal generated noise, a series of experiments on synthetic data has shown that the process is "fail soft" for gaussian noise, giving results not significantly worse than linear processing.

1. INTRODUCTION

The era of seismic arrays began in the early 1960's. The impetus was given by the inability of the current seismic technology to provide certain detection and identification of underground nuclear explosions rendering impractical a treaty to terminate the testing of underground as well as atmospheric nuclear weapons (Berkner Committee Report, 1959).

The initial recommendations of the Berkner Committee were the establishment of seismic arrays of sufficiently small spatial geometry so that high apparent velocity body waves would be enhanced at the expense of low velocity surface waves by simple summation.

Several of these small arrays were established in the U.S. and early experimentation on the signals and noise indicated that even greater improvements in signal/noise ratio could be obtained by three-dimensional filtering Burg (1964), Backus et al. (1964). These experiments led to the establishment of, first, the Large Aperture Seismic Array (LASA) in Montana and, some years later, a similar array (NORSAR) in Norway. Each of these consists of an array of subarrays. In the case of LASA there are twenty-five 7 km diameter subarrays each containing 25 seismometers. When LASA was installed it was found that at a quiet site the performance difference between linear summation and three-dimensional filtering was not great at the relatively high frequency (1 Hz) energy from distant underground nuclear explosions. This knowledge coupled with the fact that three-dimensional filtering demanded enormous amounts of digital processing led to the abandonment of three-dimensional techniques in favour of frequency filtering and linear beamforming for all except specialised processing (Seismic Array Design Handbook, 1972).

During the same period the United Kingdom Atomic Energy Authority (U.K.A.E.A.), after initial experiments on the Salisbury Plain, evolved a seismic array based on the Mills Cross (Mills & Little, 1953). This array design consists of twenty seismometers in the shape of an L, each arm of the L containing ten seismometers at a spacing of 2½ km (Birtill & Whiteway, 1965; Keen et al., 1965).

The dimensions of this type of array render the high frequency noise (> 1 Hz) incoherent for all practical purposes (Carpenter, 1965; Weichert et al., 1967) so that near optimum signal/noise ratio can be obtained by inserting delays to phase the array and then summing the sensor outputs.

One of these arrays (WRA) was installed at Tennant Creek in the Northern Territory in 1965 and is run jointly by the U.K.A.E.A. and the Australian National University.

When these arrays became operational it was found that their advantages over a single seismometer were threefold. Firstly, the inherent noise reducing capabilities allow smaller signals to be detected. Secondly, the ability to determine the three-dimensional direction of arrival of a signal allows an approximate epicentre determination to be made with one installation. Thirdly, and most importantly from the point of view of pure seismic research, a knowledge of an epicentre from the worldwide seismic network allows direct determinations of $dt/d\Delta$ (the rate of change of travel time with distance) to be made of phase arrivals. This factor coupled with the ability to separate closely spaced arrivals has proved to be a useful aid in delineating the elastic properties of the Earth's interior.

Because of the large amounts of information produced by a seismic array, dissemination of the data is normally divided into two distinct processes. The first of these is event detection in which sections of the array output containing probable events are identified. The second is event processing in which the parameters of the signal are measured.

With the U.K.A.E.A. type array the most sophisticated method of event detection is performed by initially phasing each line to a number of line velocities and then correlating combinations of these phased line sums using an integration window of about 2 s. A detection is deemed to have occurred when a

correlation output exceeds an adaptive threshold level based on the immediate past history of the noise (Weichert, et al., 1967). This detection process is carried out in the presence of seismic noise, which over short periods of time may be gaussian, but has superimposed on it other forms of noise of local origin. These other forms of noise which are normally not gaussian originate from such sources as animals, vehicles, wind gusts, rain showers, electrical storms, switching transients, etc. In addition the detection problem is compounded because the background seismic noise exhibits spatial as well as temporal variations. Under these conditions the event detection threshold is set, not to detect the maximum number of events, but to keep the number of false alarms at a manageable level.

In an effort to reduce the false alarm rate and thus increase the detection capability at WRA, Muirhead (1968a,b) suggested taking the Nth roots of the seismic outputs (with the sign preserved) before beamforming as this process enhanced coherence at the expense of amplitude. This process also had advantages over other forms of non-gaussian noise rejection, e.g., clipping, because the original data could be reconstructed by raising each sensor output to the Nth power. This meant that a one pass system requiring minimal temporary storage could be used to detect and store events.

The prospect of being better able to determine the true direction of a signal was also appreciated at that time (Muirhead 1968b, Wright & Muirhead, 1969) but was not pursued because crustal structure under WRA leading to possible plane wave deviations and loss of coherence was not known at that time.

Nth root processing has also been investigated by Kanasewich et al. (1973) who determined on synthetic as well as actual seismic data that it gave better results than any linear process that they had used. Weichert (1975), using logarithmic processing, a process designed to achieve the same non-gaussian noise suppression as Nth root processing with less computational requirements, has been able to lower the detection threshold of the YKA array by at least 0.3 m_b (approximately 6 dB) for the same false alarm rate.

The necessity for greater resolving power in azimuth and apparent velocity determinations in order to delineate seismic energy travelling through the upper 700 km of the earth's mantle has led to a closer look at the capabilities of the Nth root process for seismic array processing (Muirhead and Ram Datt, 1975). The conclusions reached in this study are that Nth root processing is only marginally worse than linear processing for both event detection and event processing if the background noise is gaussian. If the noise becomes non-gaussian the Nth root process rapidly becomes superior for lowering false alarm rates, improving signal/noise ratios and determining the direction of a signal onset.

2. THE NTH ROOT PROCESS

Let the output of the ith element of an M element phased array at time t be

$$e_i(t) = s(t) + n_i(t) \qquad (1)$$

where $s(t)$ is the coherent signal component and $n_i(t)$ the incoherent noise component.

After taking the Nth root (with the sign preserved) and summing over the whole array the normalised output is

$$E'_N(t) = \frac{1}{M} \sum_{i=1}^{M} |s(t) + n_i(t)|^{\frac{1}{N}} \text{ signum } \{s(t) + n_i(t)\} \quad (2)$$

which after raising to the Nth power becomes

$$E''_N(t) = \{E'_N(t)\}^N \cdot \text{signum } E'_N(t). \quad (3)$$

A mathematical analysis of this process has proved intractable particularly as any complete analysis should also deal with non-gaussian noise. It is difficult to define a complete set of the relevant non-gaussian noise distributions.

Two limits of this process in gaussian noise can however be defined. First, consider the case of $n_i(t) \ll s(t)$; then:

$$E''_N(t) = \{|s(t)|^{\frac{1}{N}} \text{ signum } (s(t))\}^N \cdot \text{signum } |E'_N(t)| = s(t) \quad (4)$$

so that the process passes the signal without distortion. Now consider $s(t) \ll n_i(t)$. In this case $|s(t) + n_i(t)|^{\frac{1}{N}} \cdot \text{signum } \{s(t) + n_i(t)\}$ approximates to ± 1 as N becomes large. Thus $E'(t)$ approximates to the sign process (Melton and Karr 1957) and $E'(t)$ rather than $E''(t)$ approximates to the true signal shape as M and N become large.

To determine the expected values of $E''_N(t)$ between the two extremes we have synthesised a signal in sets of gaussian noise and determined the value of $E''_N(t)$ averaged over a large number of sets for different signal-to-noise ratios and different values of N. Specifically, 10^4 sets of gaussian numbers having zero mean and unit standard deviation were generated. To each sample in each set a constant corresponding to a coherent signal was added, each set was combined according to equations (2) and (3) and the average value of $E''(t)$ for each signal level obtained. The results for four different values of N are plotted in figure 1.

As anticipated, the normalised output for N = 1 is equivalent to the input because as can be seen from equation (3) the Nth root process degenerates to linear processing for N = 1. Verification of the two extreme predictions is also indicated. For large signal/noise ratio the expected value of $E''_N(t)$ asymptotically approaches the signal input for all values of N. Similarly for small values of signal/noise ratio the expected value of $E''_N(t)$ approaches $s(t)^{\frac{1}{N}}$. This figure provides graphical evidence that the Nth root process introduces signal distortion and is of little use if the prime consideration is to preserve the signal wave shape particularly when the signal/noise ratio is small. However, as will become evident, the larger the value of N the more the background noise level is suppressed so that effectively the Nth root process enhances the signal/noise ratio where the signal is largest. This is believed to be the reason why Nth root processing is superior for determining signal directions.

3. SIGNAL DETECTION

The ability of the Nth root process to suppress non-gaussian noise is illustrated in figures 2 and 3. Figure 2 shows some of the seismometer outputs from the WRA array with a large synthetic noise spike added to one channel. The single noise spike has been introduced for clarity. The extension of the results to noise transients lasting many cycles or even incoherent transients on more than one channel is trivial. The linear processed output with the array phased to the event is shown in figure 3(a) and the Nth root processed output using a value of N = 2 in figure 3(b). Notice that in the linear processed output the noise spike is as large as the signal in the phased array sum and is still significant in the correlator output while in the Nth root output it is suppressed almost to the point of being lost in the background noise. As may be easily verified by considering a noise spike on one channel and no noise on all other channels, the amplitude reduction for this type of noise approximates to $\frac{1}{M^N}$ on the array output so that for a 20 element array and N = 2 the noise reduction is 400.

To determine the detection capabilities of the Nth root process using gaussian noise, 10^5 sets of 20 gaussian distributed numbers were generated and combined according to equations (2) and (3) for various values of N. For each value of N the maximum value of $E''_N(t)$ was chosen as a reference detection threshold. This threshold corresponds approximately to one false alarm per array beam per day, the detection being performed on the rectified output of the total array sum without any signal averaging. The experiment based on this method of detection rather than on the averaged correlator output was performed for computational expediency. However, because the transient nature of seismic signals only allows averaging over two or three cycles it is believed that the results to be given are representative of those which would be obtained if the experiment was designed to simulate the real situation as precisely as possible. Further sets of gaussian numbers were then generated with a constant, corresponding to a signal, added and once again combined according to equations (2) and (3).

By declaring a detection when the "signal + noise" output exceeded the reference threshold, the percentage detection capability for various values of N and for different signal levels was obtained and is shown in figure 4(a). This figure indicates that for a value of N = 2 the 50% detection capability is only marginally worse than the linear case and the N = 2 process could be used instead of N = 1 with small loss of detection capability in pure gaussian noise. As the value of N is increased the loss of detection capability increases but for N = 8 the 50% detection capability is only 3 dB worse than for linear processing.

This experiment was then repeated with detection thresholds equal to 1.4 and 2.0 (3 dB and 6 dB) times the reference thresholds the results of which are shown in figure 4(b) and 4(c). Figure 4(b) shows that, if the detection threshold must be raised in order to keep false alarms at an acceptable level, a point is soon reached where the detection capabilities in the presence of gaussian noise are independent of the value of N. If the detection threshold must be raised still further, the Nth root process has a better detection capability and this capability increases with increasing values of N.

Of further interest is the fact that for N = 1 the signal level for a 50% detection capability varies in the same ratio as the detection threshold while for N = 8 the detection capability is almost independent of the detection threshold. This indicates that the Nth root process can accommodate varying ambient levels of noise much better than the linear process.

The conclusions which can be drawn from the results illustrated in figures 2, 3 and 4 are that, in the presence of gaussian noise, linear detection processing performs marginally better than the Nth root process provided

N is not large. However, if the noise is not gaussian then the Nth root process is better able to deal with it both from the point of view of suppressing noise transients and of being able to operate with a higher signal detection capability in the presence of these transients. Furthermore, figure 4 indicates that the Nth root process is better able to deal with fluctuating noise levels even if the noise is gaussian.

4. SIGNAL/NOISE ENHANCEMENT

In figure 4(a) it was shown that for gaussian noise the Nth root process was marginally worse at detecting signals than linear processing. Here it is sought to evaluate the Nth root process for signal/noise enhancement in the presence of noise which is not gaussian but which has the same properties at each sensor output.

To accomplish this, six signals having the same azimuth but different amplitudes and apparent velocities were synthesized. Three representative channels containing these signals are shown in the first three traces of figure 5. To these signals uniformly distributed noise band pass filtered in the frequency range 1/2 to 4 Hz was added to each of the 20 channels. The same three channels containing both signals and noise are shown in traces 4 to 6 of figure 5. Traces 7 to 10 of this figure show the processed output of the array phased to the last signal for N = 1, 2, 4, 8. The array output was obtained by phasing and summing each line of the array according to equations (2) and (3) and then multiplying and integrating the two line sums for half a second. For N = 1 the signal is still hidden in the background noise but rises out of it as N increases. Although this artificial noise is not gaussian, its character is indistinguishable from filtered seismic noise. As such it illustrates the potential of the method for increasing signal/noise ratio and improving the clarity of a signal. This figure also shows that the Nth root process is not equivalent to raising the linear output to the Nth power.

5. AZIMUTH AND APPARENT VELOCITY DETERMINATION

The suppression of both noise and coherent signals having different velocities in figure 5 suggests that the Nth root process may also have an application in refining the determination of signal directions. To test this hypothesis, a section of an event coda from WRA was chosen as the background noise. To avoid ambiguity, the particular event chosen was a presumed underground nuclear explosion from Novaya Zemlya because characteristically explosions have simple source functions giving rise to a coherent signal onset lasting only two or three cycles. To this event coda a synthetic signal was added and the apparent velocity determined by locating peaks in a two second correlator output placed over the signal as the array was phased to different velocities. In order to improve the statistical reliability, this experiment was repeated a number of times with both the synthetic signal and the correlation window moved a further second along the records. The results are plotted in figure 6 for N = 1, 2, 4, 8. It should be remarked that because of the low signal level there were, in some cases, other maxima in the correlation output. These were present for each value of N so that in all cases the maxima nearest the correct apparent velocity was chosen. These results show that, without exception, the apparent velocity given by the Nth root process is as good as, or better than, that given by the linear process and imply that the Nth root process will give more refined determinations of secondary phases of an event. If the phase had an unknown azimuth as well as an unknown apparent velocity the search would have to be carried out in two dimensions. Computationally, this is identical to a one dimensional search, so that statistically the results will also be identical.

6. CONCLUSIONS

A non-linear method of processing the output of an array has been given. Except for amplitude information, the Nth root process is never more than marginally worse than linear processing. For non-gaussian noise, Nth root processing can give significant improvements in event detection, signal/noise enhancement and information on the direction and apparent velocity of the signal.

Although the results presented have been based on the output of a seismic array it is believed that similar results would be obtained on other types of arrays or on single sensors where the desired signal is repetitive.

7. ACKNOWLEDGEMENTS

We wish to thank Professor A.L. Hales for his help and interest during the course of this study.

REFERENCES

No.	Author	Title
1	Berkner Committee Report	"The need for fundamental research in Seismology". Report on the panel on Seismic Improvement, L.V. Berkner Chairman. Department of State. 1959
2	Backus, M., Burg, J., Baldwin, D. and Bryan, E.	"Wide-band extraction of mantle P waves from ambient noise". Geophys., 29, pp. 672-692, 1964
3	Birtill, J.W. and Whiteway, F.E.	"The application of phased arrays to the analysis of seismic body waves". Phil. Trans. A, 258, pp. 421-493, 1965
4	Burg, J.,	"Three dimensional filtering with an array of seismometers". Geophys., 29, pp. 693-713, 1964
5	Carpenter, E.W.	"An historical review of seismometer array development". Proc. I.E.E.E. Vol. 53, No.12, pp. 1816-1821, 1965
6	Kanasewich, E.R., Hemmings, C.D. and Alpaslan, T.	"Nth-root stack non-linear multichannel filter". Geophys., 38, pp. 327-338, 1973
7	Keen, C.G. Montgomery, J., Mowat, M.W., Mullard, J.E. and Platt, D.C.	"British seismometer array recording systems". The Radio and Electronic Engineer, 30, pp. 297-306, 1965
8	Melton, B.S. and Karr, P.R.	"Revealing signal coherence". Geophys., 22, pp. 553-564, 1957
9	Muirhead, K.J.	"Eliminating false alarms when detecting seismic events automaticaliy". Nature, 186, pp. 704, 1968 a
10	Muirhead, K.J.	"The reduction and analysis of seismic data using digital computers". Ph.D. Thesis, University of Tasmania, 1968 b
11	Muirhead, K.J. and Ram Datt	"The Nth root process applied to seismic data". (In preparation), 1975
12	Seismic Array Design Handbook	ARPA Order No. 800, Prepared by I.B.M. Federal Systems Division, Gaithenburg, Maryland, 1972
13	Weichert, D.H., Manchee, E.B., and Whitham, K.	"Digital Experiments at twice real time speed on the capabilities of the Yellowknife seismic array". Geophys. J., 13, pp. 277-295, 1967

No.	Author	Title
14	Weichert, D.H.	"Reduced false alarm rates in seismic array detection by non-linear beamforming". Geophysical Research Letters, Vol.2, No.4, pp. 121-123, 1975
15	Wright, C. and Muirhead, K.J.	"Longitudional waves from the Novaya Zemlya explosion of October 27th, 1966 recorded at the Warramunga Seismic Array". J. Geophys. Res., 74, pp. 2034-2048, 1969

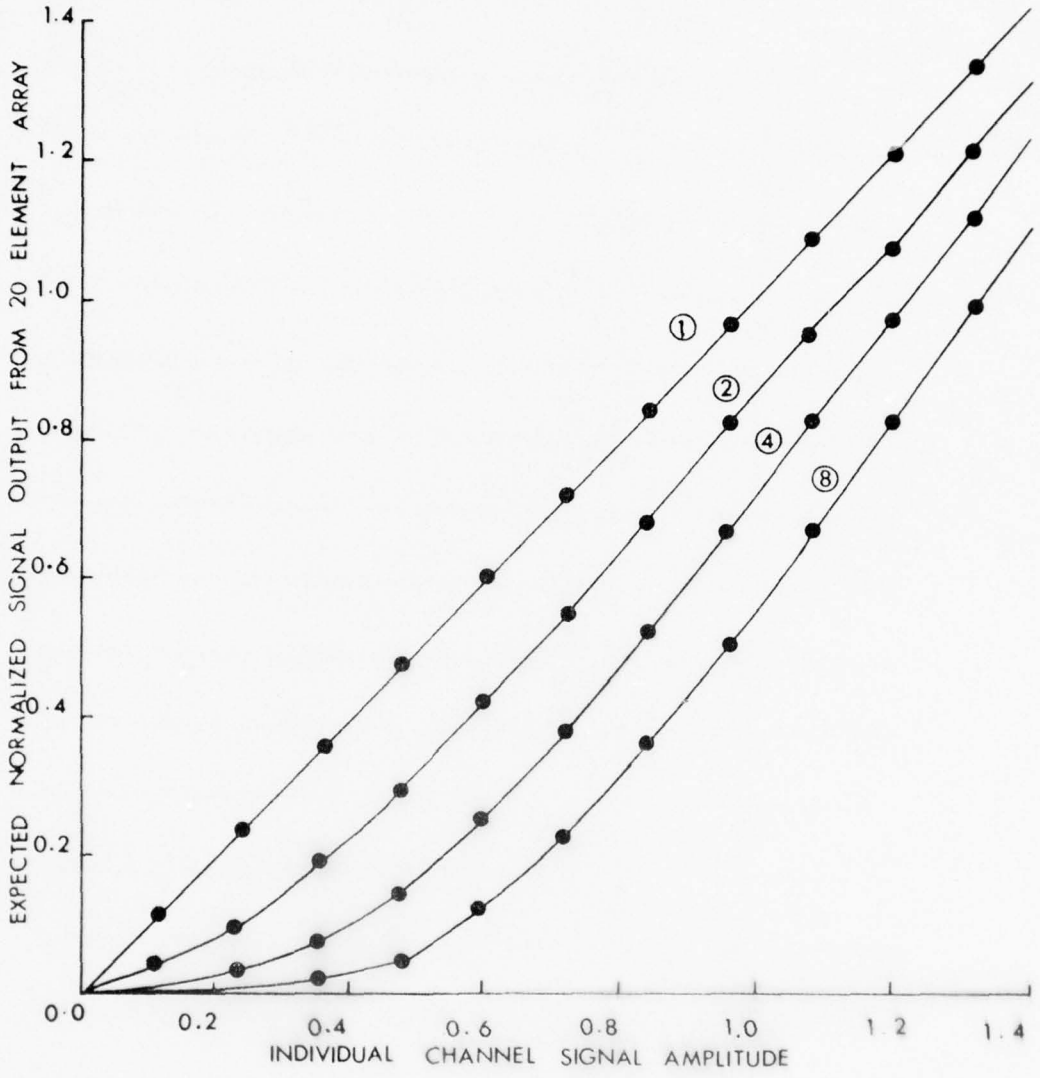


Figure 1. DIAGRAM ILLUSTRATING THE EXPECTED ARRAY OUTPUT FOR DIFFERENT VALUES OF N AS A FUNCTION OF SIGNAL/NOISE RATIO. The noise is Gaussian with zero mean and unit standard deviation.

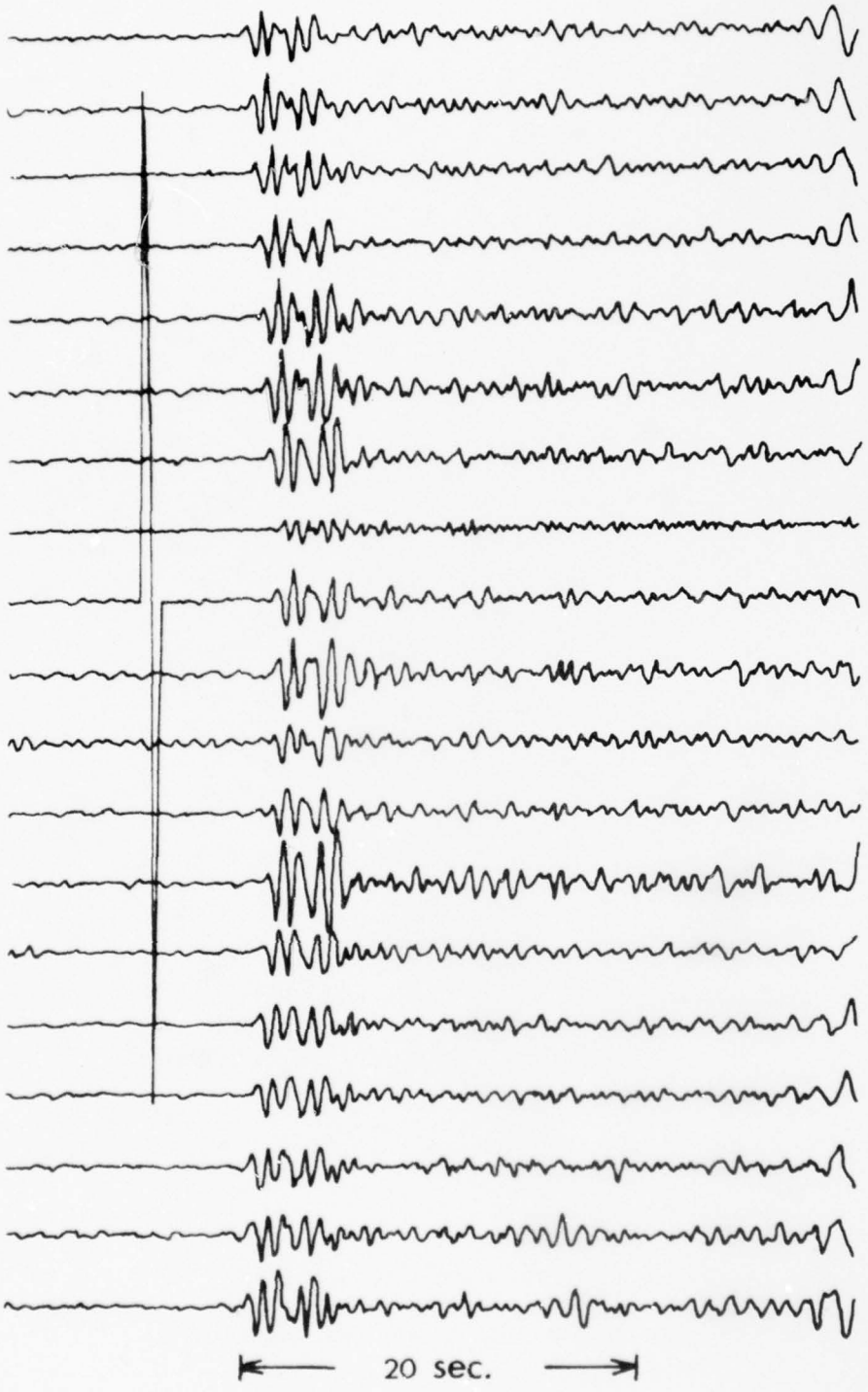


Figure 2. SAMPLE OUTPUT FROM WRA SHOWING A SMALL EVENT ONSET AND A LARGE SYNTHETIC NOISE SPIKE ADDED TO ONE CHANNEL.

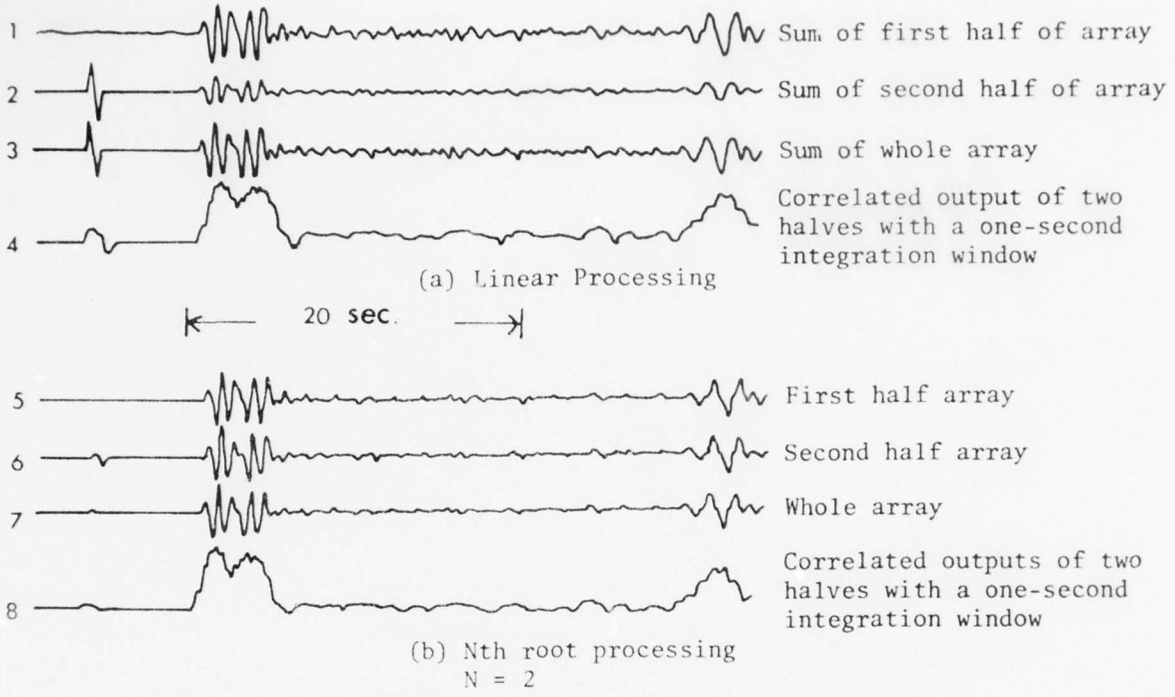


Figure 3. PROCESSED OUTPUT OF EVENT SHOWN IN FIGURE 2.

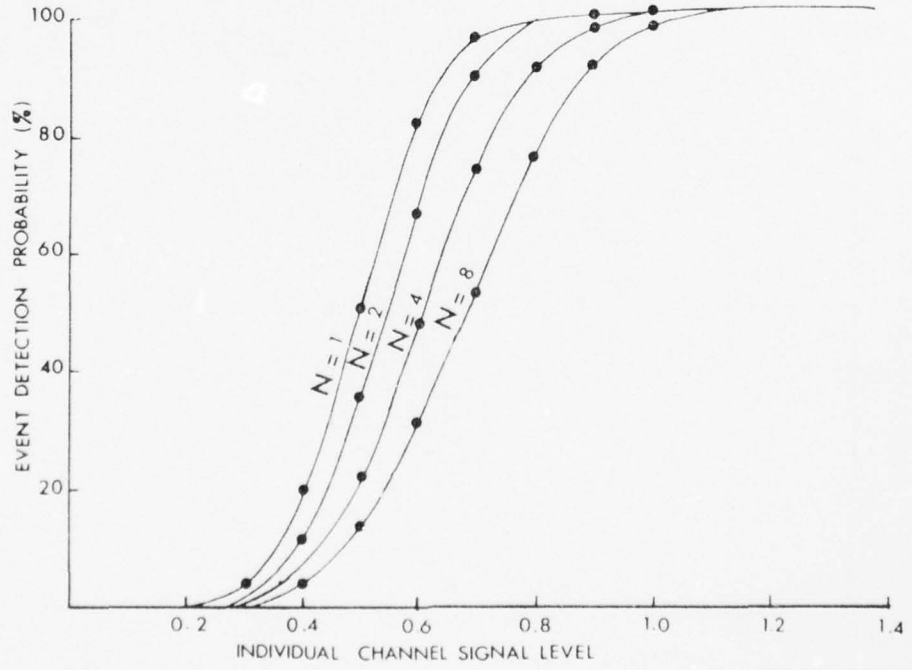


Figure 4(a). PROBABILITY OF EVENT DETECTION USING A DETECTION THRESHOLD EQUAL TO THE REFERENCE THRESHOLD.

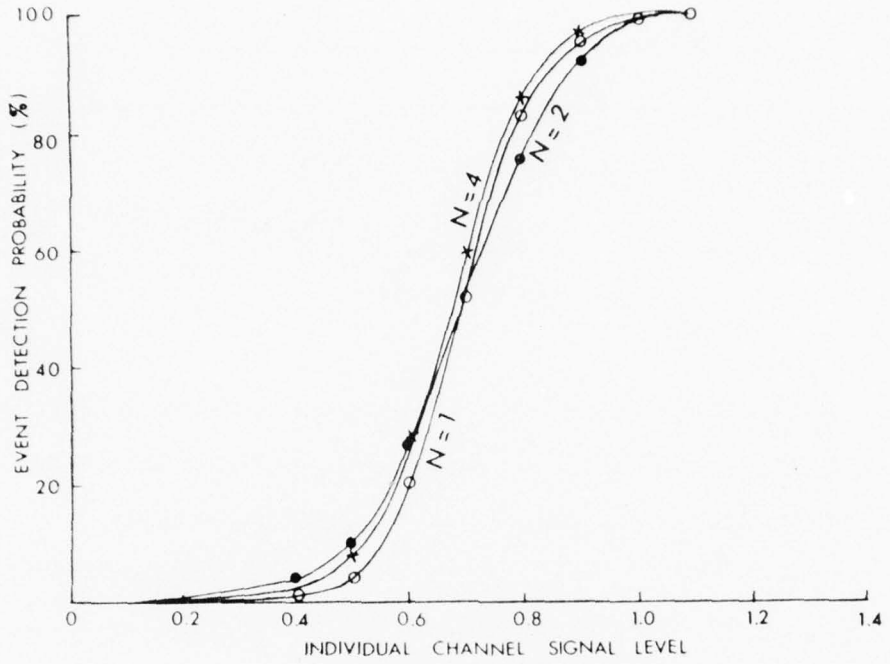


Figure 4(b). PROBABILITY OF EVENT DETECTION USING A DETECTION THRESHOLD EQUAL TO 1.4 TIMES THE REFERENCE THRESHOLD

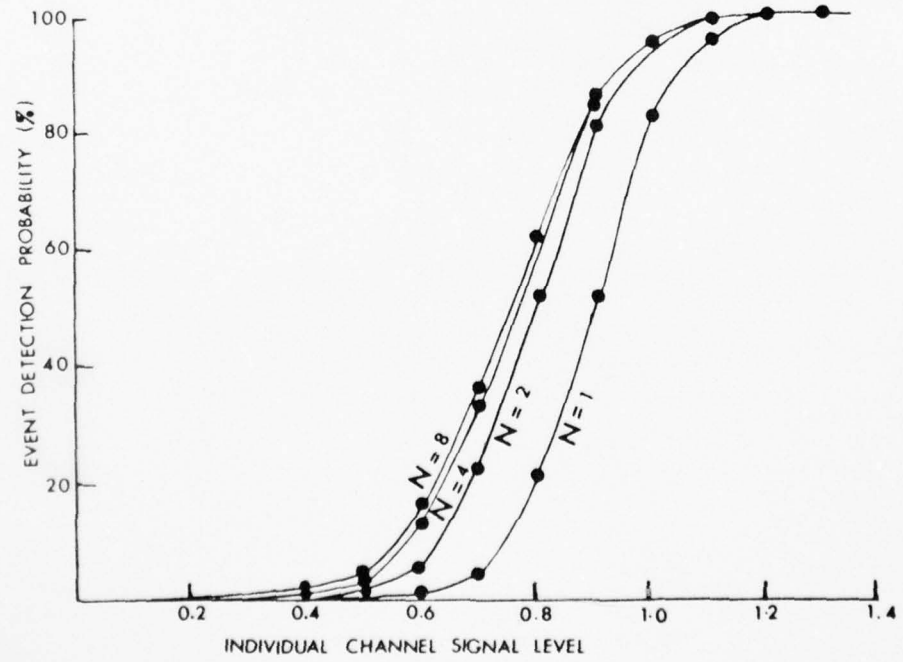


Figure 4(c). PROBABILITY OF EVENT DETECTION USING A DETECTION THRESHOLD EQUAL TO TWICE THE REFERENCE THRESHOLD

Figure 5. DIAGRAM ILLUSTRATING THE SIGNAL/NOISE IMPROVEMENT OF THE NTH ROOT PROCESS. Traces 1 to 3 show three synthesised sensor outputs containing six phase arrivals. The phase velocities are given in kilometres/second. Traces 4 to 6 show three sensor outputs with filtered noise added. Traces 7 to 10 show the array output phased to the 74 km/s signal with N = 1, 2, 4, and 8 respectively.

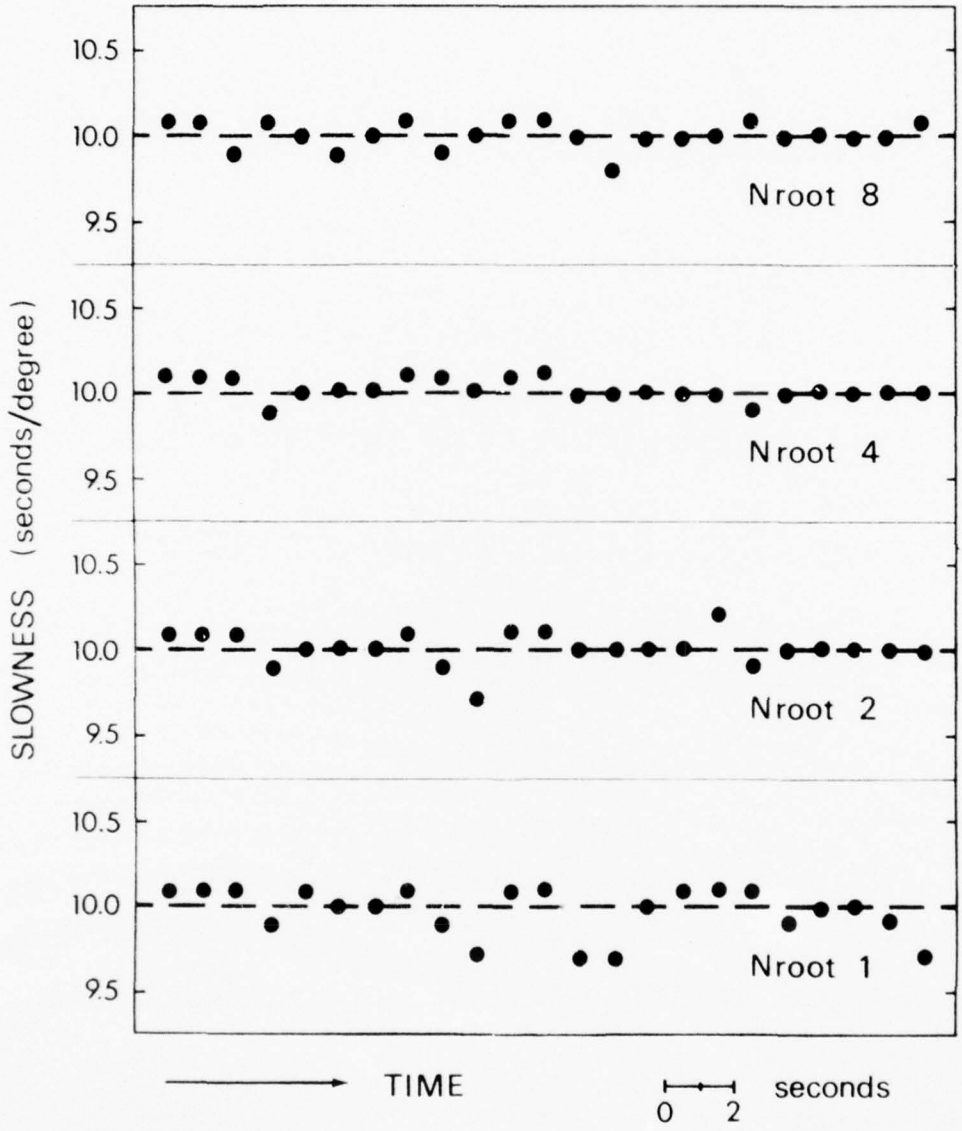


Figure 6. RESULTS OF APPARENT VELOCITY DETERMINATION OF A SYNTHETIC SIGNAL PLACED IN THE CODA OF AN EVENT USING THE NTH ROOT PROCESS. The synthetic signal was added to the noise with a slowness (reciprocal of apparent velocity) of 10 s/degree.

MEASUREMENT OF PHASE VELOCITY OF MICROSEISMS, USING A SMALL EXPANDING SEISMIC ARRAY

M.W. Asten, School of Earth Sciences,
Macquarie University

SUMMARY

An experiment designed to measure the phase velocity dispersion of surface (Rayleigh) waves has been made in the Sydney Basin. An array of stations large enough to adequately resolve long wavelengths, and with station density sufficient to suppress wavenumber aliasing is not logistically feasible. However, the properties of such an array may be partially synthesized using a three station array whose diameter is progressively increased, using procedures to minimize the number of station relocations.

Due to rapid variation in the apparent direction of wave propagation, a digital processing technique to scan large amounts of data, for segments showing maximum coherency, was used. By assuming plane wave propagation, meaningful velocity estimates were possible for waves of 5 to 0.15 Hz. Some evidence exists for the presence of two modes of surface waves, but further interpretation is not possible.

Theoretical studies of the response of three, four and five station arrays, to plane waves from multiple sources, suggests that phase velocities from such sources may be successfully measured using an expanding array of five stations or more. Such measurements, however, are subject to bias of up to 30%, due to properties of the array beam response.

Further observations, using additional seismic stations, are required before geological information can be inferred from the velocity dispersion data; procedures to use up to seven stations, without sacrificing field mobility, are the basis of continuing study.

1. INTRODUCTION

Microseisms are ambient ground oscillations occurring within the general frequency range 0.01 Hz to 30 Hz. The band 0.1 to 10 Hz is of interest to the exploration geophysicist in that at least part of the microseismic noise propagates as a wavemotion, the mode, direction and velocity of which may be established by observations with a suitably designed array of seismometers.

The wavemotion propagation parameters carry information characteristic of the elastic parameters of the earth's crust, to depths of the order of 20 km. Changes in deduced elastic parameters are relatable to rock types and thus may be used to infer depth and extent of sedimentary basins. It is the purpose of this study to consider what type of array and data processing may be required for microseismic studies of value to an exploration geophysicist.

2. MICROSEISM CHARACTERISTICS

For this study, three source-types of ground oscillation (other than that generated by earthquakes) are considered:

(1) Artificially generated oscillation, e.g., by road and rail traffic, or by industrial machinery.

(2) Naturally generated local, or non-propagating oscillations, e.g., by wind pressure on vegetation and topography, or pressure variations induced by non-horizontal movement of air masses.

(3) Naturally generated propagating wavemotion originating at distant sources, such as at coast-lines by wave action, or by atmospheric cyclonic disturbances over continent or ocean.

The most useful propagation modes for this study are Rayleigh waves, which propagate along the earth-air interface with elliptical particle motion and an amplitude decaying with depth. In a layered earth where elastic parameters vary with depth, dispersion of phase and group velocities, with respect to frequency, is characteristic.

Depths of interest in studies of sedimentary basins range from 0.1 to 20 km which, assuming typical fundamental Rayleigh wave velocities in the range 1 to 3 km/s, correspond to a frequency band 10 Hz to 0.15 Hz. From previous studies of microseisms, it is possible to identify which of the three signal categories are of most relevance.

The contribution of category (1) signals is difficult to estimate prior to a survey. Observations by Sanford et al (1968) and Aki (1957) show the signals to have frequencies in the range 1 to 30 Hz and to propagate in Rayleigh or Love modes. Such signals are potentially useful in determining elastic parameters of the upper kilometre of rock strata but the direction and intensity of sources is highly non-stationary and must complicate observation techniques. In addition, sources may be local, thus making the theory of array response to a plane wave inapplicable.

Category (2) ground oscillation is regarded as noise, and can be reduced by choice of seismometer sites in areas of minimal vegetation. Oscillation due to loading effects of air currents on the earth's surface, as distinct from vegetation, has been found by Ziolkowski (1974) to lie chiefly in the frequency band 0.05 to 0.01 Hz, which is beyond the band of interest in this study.

The third category of microseismic signals is likely to prove easiest to observe.

Using the Montana Large Aperture Seismic Array, Toksoz and Lacoss (1968), Capon (1969) and Lacoss et al (1969) found predominantly compressional wave propagation at 0.5 Hz, higher-mode Rayleigh wave propagation from 0.3 to 0.2 Hz,

and fundamental mode Rayleigh wave propagation at lower frequencies. In an earlier survey, using seven seismometers in an L-shaped array of arm length 1.5 km, Toksoz (1964) measured phase velocities corresponding to fundamental mode Rayleigh waves for the frequency range 0.5 to 0.16 Hz. In the same paper, results are quoted from Archambeau et al (1963), who observed strong higher mode Rayleigh waves at frequencies above 1 Hz.

An additional difficulty in phase velocity measurements is multidirectional propagation, with the dominant direction changing by up to 90° in minutes.

In summary, Rayleigh waves are likely to exist over at least part of the frequency band of interest but their detection requires an array, and signal processing system capable of separating mixed Rayleigh mode, and compressional wave, and propagation from multiple, nonstationary sources.

3. ARRAY DESIGN

The range of wavelengths of interest, from 0.1 to 20 km, requires 17 stations for the sparsest possible array (figure 1) which fulfills the two chief requirements of (i) avoiding wavenumber aliasing at high frequencies, and (ii) having an aperture greater than the longest wavelength of interest. At this experimental stage, an array of such size is not logistically viable as an exploration tool, particularly if three component seismometers are desired. Procedures for using a three, four, five or seven station array, and progressively expanding it with a minimum of station relocations, have been developed.

A three-station array of seismometers 1, 2 and 3 is constructed at stations 1, 2 and 3 of figure 2 and called "Array 1". After recording suitable data, seismometer 1 is moved to station 4, to form "Array 2", and recording is continued. Seismometer 2 is then moved to station 5 and so on. The seismometers are thus moved out progressively along the arrowed paths, always maintaining the right-angle triangle array geometry. A four-seismometer array may be designed similarly, by having a fourth seismometer permanently installed at station 0.

For a five seismometer array, (figure 3), seismometers 0 to 4 are initially installed at stations 0 to 4. After recording, seismometers 1 and 2 move to stations 5 and 6 to form "Array 2", thus expanding the array with only two seismometer relocations. As with the three seismometer array, the geometry is preserved, and step by step expansion can be repeated until the desired range of array sizes has been covered.

The station configuration of figure 3 may also be used for an expanding seven or nine seismometer array.

4. ARRAY RESOLVING POWER

Properties of the "expanding array" have been examined, using synthetic data. A wave source is considered to be at infinite distance, and waves are given constant amplitude across the array. For multiple sources, the amplitude at each of N stations, is computed from a linear sum of the influence of M sources;

$$A_n = \sum_{j=1}^{M} Z_j \exp\{i2\pi(ft + \underset{\sim}{k}_j \cdot \underset{\sim}{r}_n)\}$$

where

A_n = complex amplitude at station n

$\underline{r}_n$ = displacement of station n from array origin

k_j, Z_j = radian wavenumber, amplitude of plane waves of frequency f, from the j^{th} source

i.e., $\underline{k}_j = -\underline{v}/f$, where $\underline{v}$ = phase velocity

From the set A_n, (n = 1, N), a coherency matrix is computed, which is subjected to a conventional two-dimensional wavenumber transform on a 41 by 41 point grid. A plane wave is thus represented as a vector in wavenumber space, oriented towards the source, as described by Capon (1969).

The reasoning behind the expanding array procedure is illustrated, for the case of a single, stationary source, in figures 4 to 7. Since lobe size of the array beam pattern is inversely proportional to array size, only an approximate vector wavenumber estimation is possible from data acquired by smaller arrays. However, this estimate enables more precise estimation of wavenumber with larger arrays, despite the presence of spurious peaks on the wavenumber plot, due to wavenumber aliasing. The procedure should remain viable if the azimuth varies over less than a quadrant.

When multiple sources are considered, the limitations of the small array are more apparent. The Rayleigh criterion for resolution of two sources separated by an azimuth of 120° is satisfied (figure 8) by the five station "Array 3" configuration. However, when the two sources are closer together on the wavenumber map (figure 9), resolution is only possible if a correspondingly larger array is used, which then may bring confusion in the form of wavenumber aliasing (figure 10).

A serious difficulty evident with wavenumber plots, when multiple sources are not resolvable, is illustrated in figures 11 to 13. In all cases tested to date two or more unresolved, or partially resolved, peaks on a wavenumber plot separated in azimuth by 90° or less, appear as a single peak biased towards a mean wavenumber. Thus multidirectional plane waves with the same velocity modulus appear on the wavenumber plot as a single diffuse peak, biased towards a shorter wavenumber, i.e., a higher velocity. If sources are separated in azimuth by more than 90°, a bias towards higher or lower velocities is possible, depending on azimuths relative to the array geometry (compare figures 13 and 14). Typical bias errors in wavenumber estimations of sources incompletely resolved by the five station array are of the order 10% to 15%, with a maximum bias observed, in these studies, of 30%.

A four station expanding array may be constructed using the configuration of figure 2, with addition of a fixed station at the array origin. The reduced spatial sampling of this array, compared with a 5 station array, results in larger biasing errors of wavenumber estimates of inadequately resolved multiple sources. The increased tendency to wavenumber aliasing also reduces reliability of wavenumber estimates, because the direction of biasing is more difficult to predict when sources have wavenumber differences comparable with separation of side-lobes of the array beam pattern. Figures 15 and 16 provide an illustration; the array response to a single source only (figure 15) has a side-lobe peak near the wavenumber of the second source in figure 16, with a resultant array response showing the two peaks biased to wavenumber moduli 20% low and 35% high, respectively.

The usefulness of the simple three station array is limited to situations where a single source only is dominant, because no redundancy of spatial information exists to give any indication when multiple sources exist. The side-lobe structure of the response of such an array makes identification of the main lobe an exceedingly hazardous task, unless the source direction is known (figure 17).

Limitations of the arrays when waves propagate from a near source have been

considered only superficially. To a first approximation, waves of velocity v from a point source distant R from an array of radius r, appear to the array, when plane wave theory is used, to emanate at velocity

$$v' = v/\cos\theta, \quad \theta = \arctan(r/R),$$

from a diffuse direction. Velocity estimates of waves from a source at distance 3r are therefore expected to be biased 5% high.

Summarizing the theoretical limitations of the arrays considered, the five station expanding array appears capable of resolving multiple sources with significant, but partially predictable, bias of wavenumber estimates of incompletely resolved sources. The four station array can resolve multiple sources but may give heavily biased wavenumber estimates. The three station array is of little value unless the wave source is single and of known azimuth.

5. PROCESSING OF FIELD DATA

A trial survey designed to test instruments and field procedure has been conducted in the Sydney Basin, 70 km WSW of Sydney. Three stations only were used, in array configurations 1 to 9 of figure 2. Signals from buried, 0.7 Hz vertical Willmore seismometers were amplified 90 or 96 db, with pass band 0.1 - 10 Hz, at each site and carried in analogue form on twin-flex to a truck-mounted Interdata 70 minicomputer at station 7. The need for cable links to each station limited maximum array size to a 2 km diameter, or a tenth of the desirable maximum. At the computer, data was again filtered with a low pass of 10 Hz or 2 Hz, to attenuate noise pick-up in the cables. After analog to digital conversion of each of the three signals, data was multiplexed and stored on magnetic tape in files of 2048 points per channel, sampled at 10, 20, 50 or 100 ms intervals.

Data from each channel was detrended, reduced to zero mean, transformed to the frequency domain without prewhitening, and corrected for seismometer and filter transfer response. Coherencies were then computed between each pair of stations, using smoothing over a square window, spaced at intervals of 0.1 decade in frequency, and with mutual overlap of 0.75. The frequency range thus covered was from $10f_1$ to $0.4f_n$, where f_1, f_n are fundamental and Nyquist frequencies, respectively. Coherencies ranged between 0.8 and 1.0 for frequencies below 0.5 Hz, but dropped rapidly at high frequencies. This feature is partially attributable to the likelihood of non-propagating seismic noise levels, or variations in source strength and direction, to occur over the duration (20 to 200 s) of a single record.

The need to locate data lengths containing high coherency signals attributable to a single source, led to development of an alternative processing procedure. Data from each file was split into 31 overlapping segments of 128 points each, with each segment being processed to obtain coherencies as described previously. The files were then split into 15 segments of 256 points, 7 segments of 512 points, and 3 segments of 1,024 points with coherencies again being computed for each segment so considered. For each of the tenthdecade frequency bands, the best 10 coherency sets obtained were saved and the remainder discarded. By repeating the entire process for all data files acquired with the same array, it became possible to locate segments of data with coherency sets averaging 0.9, 0.8 and 0.7, in frequency ranges 1 to 2 Hz, 2 to 3 Hz and 3 to 5 Hz, respectively.

Having obtained the highest possible coherencies from segments of data, the coherency phases were used to compute direction and velocity of the plane wave assumed responsible for the phase relations. Thus for each of the eight array configurations, up to ten velocity estimates at each tenth-decade frequency were

possible, making a maximum of 80 velocity estimates at each frequency. The number of estimates is progressively lower for frequencies above 1 Hz, due to the difficulty of estimating velocities in the presence of wavenumber aliasing with the larger arrays. At low frequencies below 0.3 Hz, the number is also reduced since spectral estimates required use of the full data length of each file. The number of estimates at these low frequencies was augmented by processing data of lengths up to 200 s, acquired with the largest three arrays.

Vector velocity estimates at each frequency were plotted in polar coordinates to obtain graphs of the form of figure 18. At each frequency, estimates of apparent source direction from arrays 1 and 2 cluster towards the north of NNE, and cluster progressively further towards south-east for the larger arrays. This change in azimuth of approach of microseisms is observable between 28/11/74 and 1/12/74 on the Riverview seismic observatory three-component records, and is therefore representative of a real movement of a principal microseismic source. The remaining scatter of velocity and azimuth estimates is a function of statistical uncertainty of coherency phase estimates, and theoretical limitations of the three station array already discussed.

Clustering of velocity moduli at each frequency is illustrated by plotting histograms of number of estimates vs apparent phase velocity (figure 19). The set of histograms show four features:

(i) Frequencies 1.5 to 2 Hz show a maximum number of velocity estimates to be above 7 km/s. These could correspond to a predominantly compressional mode of wave propagation, or could be the result of extreme biasing due to multiple sources of comparable strength.

(ii) Frequencies below 0.25 Hz show decreasing velocity, the most likely explanation of which is dominance of nonpropagating seismic noise due to atmospheric loading effects.

(iii) Apart from features (i) and (ii), there appears to be a frequency dispersion of apparent velocity estimates.

(iv) A bifurcation of histogram peaks occurs at frequencies above 2 Hz. With (iii) above, this is consistent with two or more modes of Rayleigh wave propagation.

6. CONCLUSIONS

The quality of results to date, summarised in figure 19, does not warrant further interpretation, but in view of the partial successes, despite the theoretical limitations of the three station array, and the limitation on array sizes used, further observations are planned. To obtain arrays of diameter up to 10 km, costly telemetry links are required. This makes five seismometers the maximum logistically possible in the foreseeable future, although cable linkage to additional seismometers is possible with small arrays. Use of three-component seismometers is indicated, to enable particle motion studies to be used in identification of wave propagation modes.

Conventional frequency wavenumber analysis for such an array gives biased wavenumber estimates in the presence of multiple sources. Interference due to such multiple sources can be reduced by selecting short data lengths, but at the cost of greater statistical uncertainty of phase relations across the array, and therefore of vector velocity estimates. A study of effects of wavenumber window heights, particularly those for high resolution wavenumber analysis (Capon, 1969) is required, to enable maximum return of unbiased wavenumber estimates from the very limited spatial sampling of data planned.

REFERENCES

No.	Author	Title
1	Aki, K.	"Space and Time Spectra of Stationary Stochastic Waves, with Special Reference to Microtremors." Bull. Earthq. Res. Inst. 35 (3): 415-456 1957.
2	Archambeau, C.B., Alexander S.S. and Bieler, E.	"Identification and Elimination of Micro-seismic Noise at Depth using Theoretical Rayleigh Wave and Observed Noise Displacement" Report, Contract AF 33(600)42890, United Electrodynamics, Pasadena, Calif. 1963.
3	Capon, J.	"High Resolution Frequency-Wavenumber Spectrum Analysis" Proc. IEEE 57 (8) : 1408 - 18, 1969.
4	Lacoss, R.T., Kelly, E.J. and Toksoz, M.N.	"Estimation of Seismic Noise Structure using Arrays" Geophysics 34 (1) : 21-38, 1969.
5	Sanford, A.R., Carapetian, A.G. and Long, L.T.	"High Frequency Microseism from a Known Source" Bull. Seism. Soc. Am. 58 (1) : 325-38, 1968.
6	Toksoz, M.N.	"Microseisms and an Attempted Application to Exploration" Geophysics 24 (2) : 154-177, 1964.
7	Toksoz, M.N. and Lacoss, R.T.	"Microseisms : Mode Structures and Sources" Science 159 : 872-3, 1968
8	Ziolkowski, A.	"Prediction and Suppression of Long Period Non-Propagating Seismic Noise" Bull. Seism. Soc. Am. 63 (3) : 937-58, 1973.

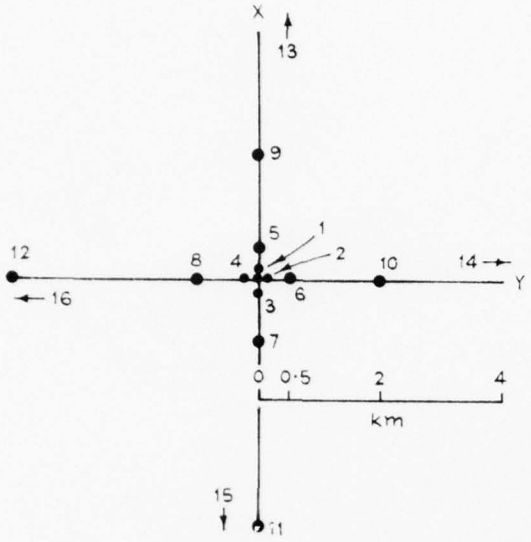


FIGURE 1. SPARSEST POSSIBLE ARRAY FOR OBSERVING WAVELENGTHS 0.1-20 km (Stations 13, 14 are 8 km from origin, and stations 15, 16, are 16 km from origin)

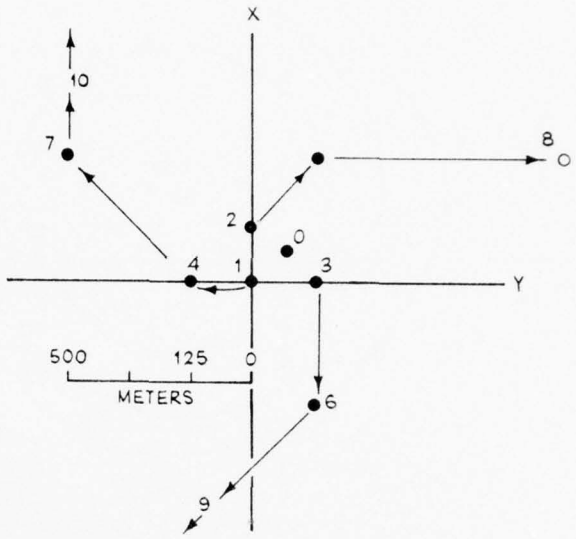


FIGURE 2. A 3-SEISMOMETER EXPANDING ARRAY (Stations 1, 2, 3 form array 1; stations 2, 3, 4 array 2; stations 3, 4, 5 array 3; etc.)

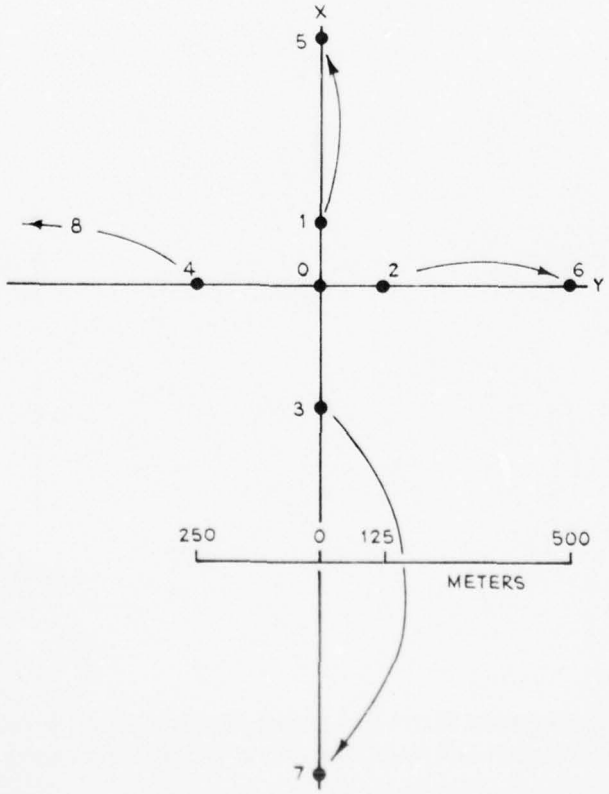


FIGURE 3. A 5-SEISMOMETER EXPANDING ARRAY (Stations 0, 1, 2, 3, 4 form array 1; 0, 3, 4, 5 6 form array 2; 0, 5, 6, 7, 8 form array 3; etc.)

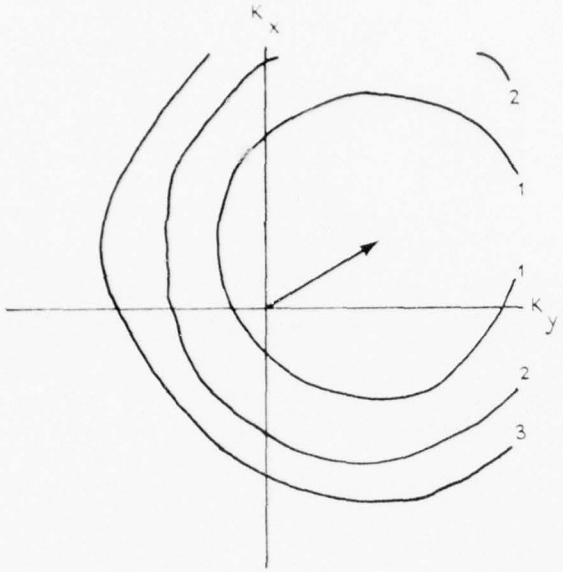


FIGURE 4

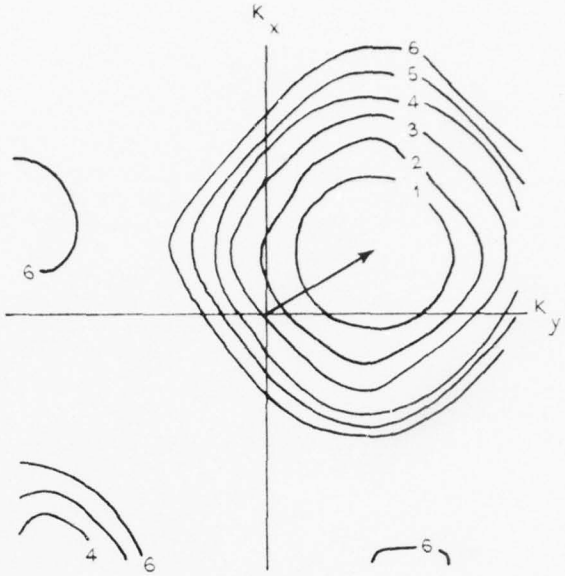


FIGURE 5

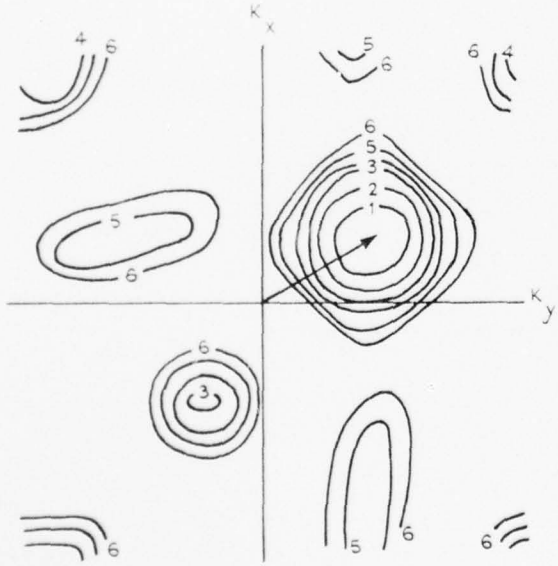


FIGURE 6

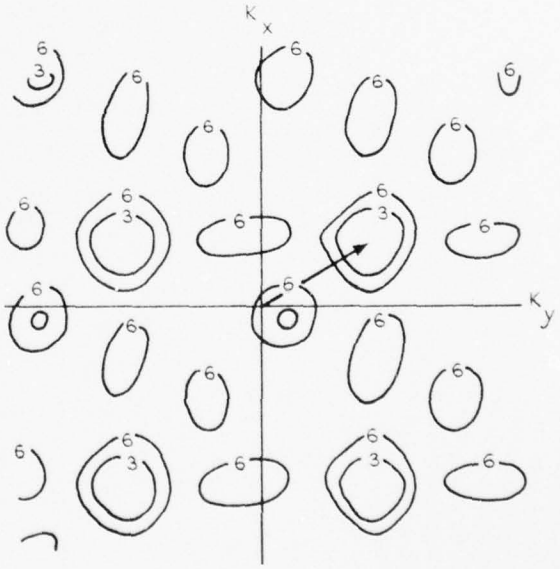


FIGURE 7

FIGURES 4 - 7. WAVENUMBER PLOTS FOR RESPONSE OF 5-SEISMOMETER ARRAYS 1-4, TO A SINGLE, 1 Hz, 2 km/s PLANE WAVE (Contours in dB relative to maximum)

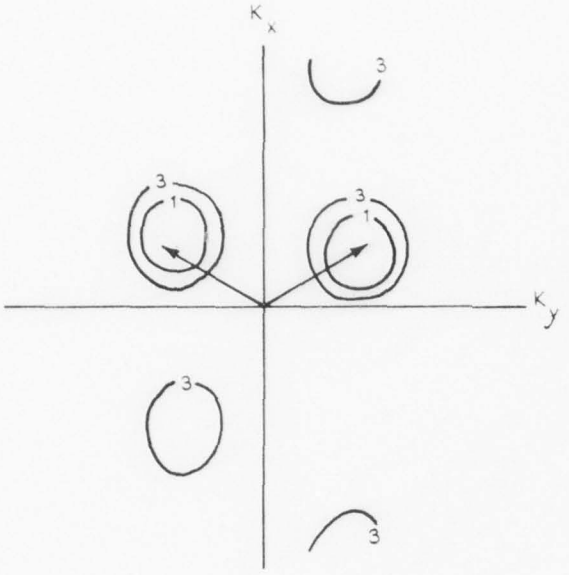


FIGURE 8

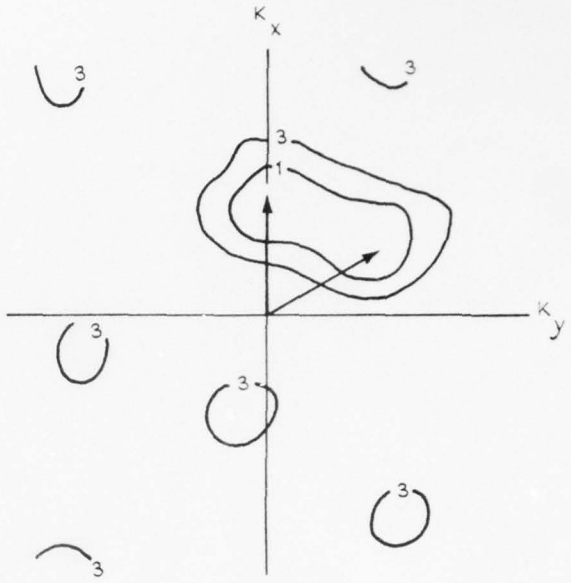


FIGURE 9

FIGURES 8 - 9. WAVENUMBER RESOLUTION OF 1 Hz, 2 km/s, WAVE FROM 2 SOURCES, USING ARRAY 3

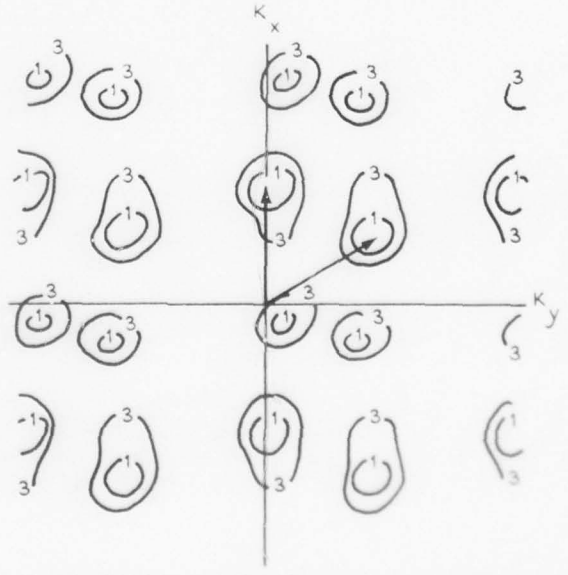


FIGURE 10. WAVENUMBERS SHOWN IN FIGURE 9 MAY BE RESOLVED USING ARRAY 4

AD-A047 017 WEAPONS RESEARCH ESTABLISHMENT SALISBURY (AUSTRALIA) F/G 12/1
PROCEEDINGS OF SYMPOSIUM SIGNAL PROCESSING FOR ARRAYS,(U)
MAY 77 H A D'ASSUMPCAO

UNCLASSIFIED WRE-MISC-1 NL

3 OF 3
ADA
047017

END
DATE
FILMED
1-78
DDC

Only the polarity of each sample is retained, and transmitted over a short telemetry link to decouple SCAR and ORB electrical grounds. Each pair of polarity bits is essentially a phase-quadrant descriptor. The total record for each 0.5 ms transmission is approximately 0.5 s, 2048 complex samples from each hydrophone.

Beamforming is accomplished by adding together the hydrophone outputs at the time delays and phase changes appropriate for the particular angular beam and range. The formulas for the time delays were developed to allow rapid calculation using a micro programmed Microdata 3200 in conjunction with the main PDP 11 minicomputer.

Consider the transmitter as the origin of a coordinate system and distances to be given in sound-seconds. A "beam" corresponds to a direction given by a unit vector, $\vec{u}$, and each hydrophone is specified by a position vector, $\vec{p}$. For each beam and hydrophone two constants are stored in memory (32768 pairs):

$$a = \vec{p} \cdot \vec{u}, \tag{1}$$

$$b = |\vec{p}|^2 - a^2. \tag{2}$$

In a moment it will become clear that it is convenient to define a function

$$y(x;\ a,b) = \sqrt{(x - a)^2 + b} - (x - a). \tag{3}$$

The first focusing law is based on a point reflector at position $\vec{ru}$. The total distance from the transmitter to the reflector and back to the hydrophone is

$$r + |\vec{p} - \vec{ru}|,$$

where

$$r + |\vec{p} - ru| = 2r - a + y(r;\ a,b). \tag{4}$$

The second focusing law is based on a reflector at position $\vec{ru}$. The total distance from the transmitter to the reflector and back to the hydrophone is

$$|\vec{p} - 2\vec{ru}| = 2r - a + y(2r;\ a,b). \tag{5}$$

Equations (4) and (5) are of the same style, and are easily implemented. The term 2r is the distance (time) from array centre to the reflector and back. The term a is the correction that would be used if the reflector were so far away that the reflection could be considered a plane wave at the array. The terms y(r) and y(2r) are the spherical wave near-field corrections. This style of equation is nice to implement because the 2r term is the same for all phones, the a term is fixed parameter for each phone but range independent, and the near-field correction function is delegated to the microprocessor.

The beamforming proceeds with focusing law being the slowest change, beam direction next, and range change the innermost. At each step the application of equations (4) or (5) defines a time for each hydrophone. The complex sample closest to this is fetched from memory, and the difference between calculated time and actual sample time used as a phase correction to the closest quadrant. The real and imaginary bits from all hydrophones are summed in hardware counters. The magnitude of the resultant complex number is encoded into a 4 bit intensity

value and stored on magnetic disc for display. There is a total of 524,288 intensities for each transmitted pulse (2 laws x 256 beams x 1024 range steps).

3. CONCLUSION

This array represents a unique application of acoustic sampling to observe oceanographic processes. It has the capability to sample length and time scales, which cannot be sampled by the more conventional direct sensing instrumentation presently used by oceanographers. The goals of the experiment are: to observe the horizontal extent and temporal behavior of water density features; to observe internal wave fields and their interaction with the water density field; to describe the distribution of large (acoustically) biological point scatterers; and finally, to understand how sound is scattered by water density features.

WRE-MISC-1
Figure 1

Figure 1. Array construction

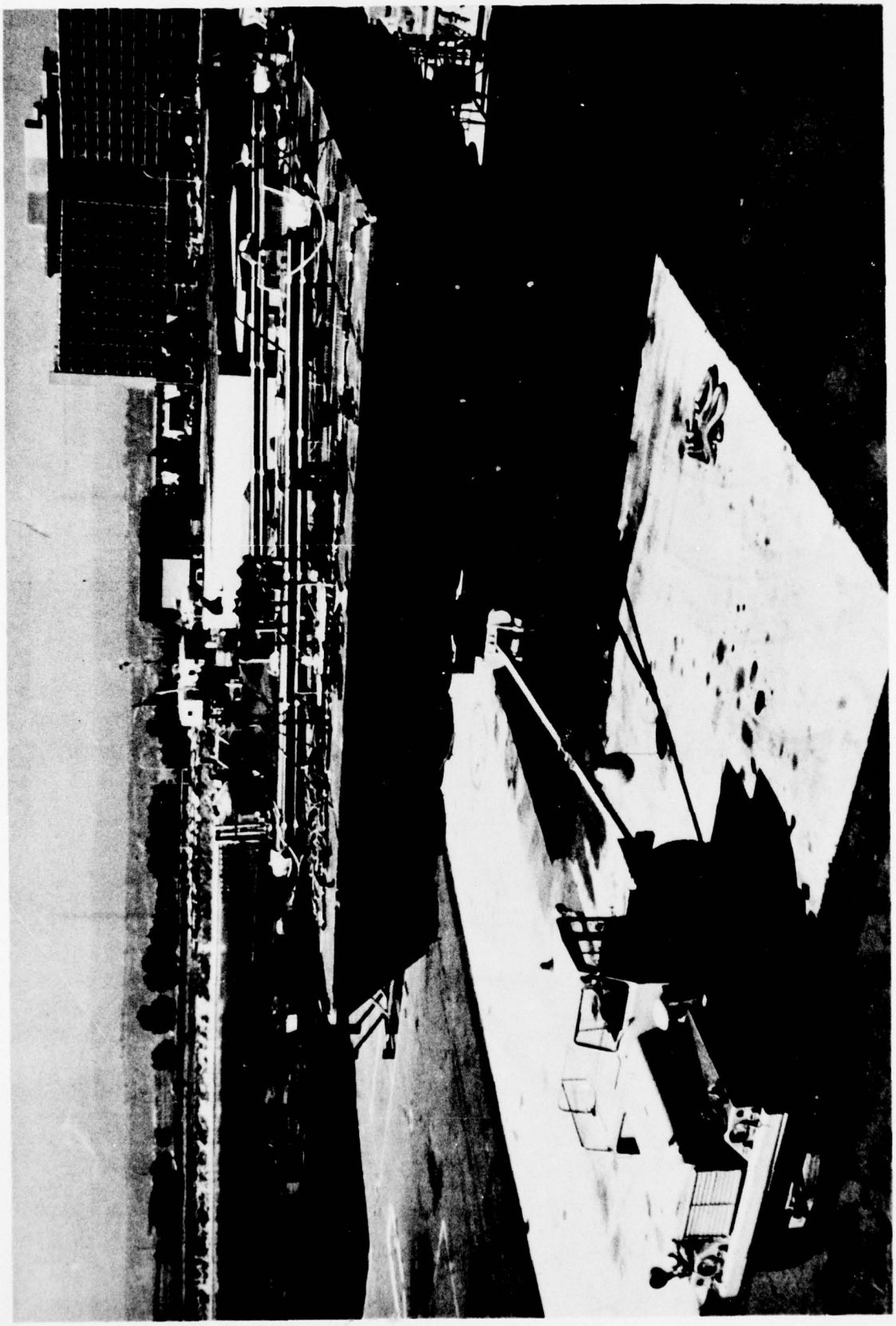


Figure 2. Fully assembled array

Figure 3. View of bottom of barge

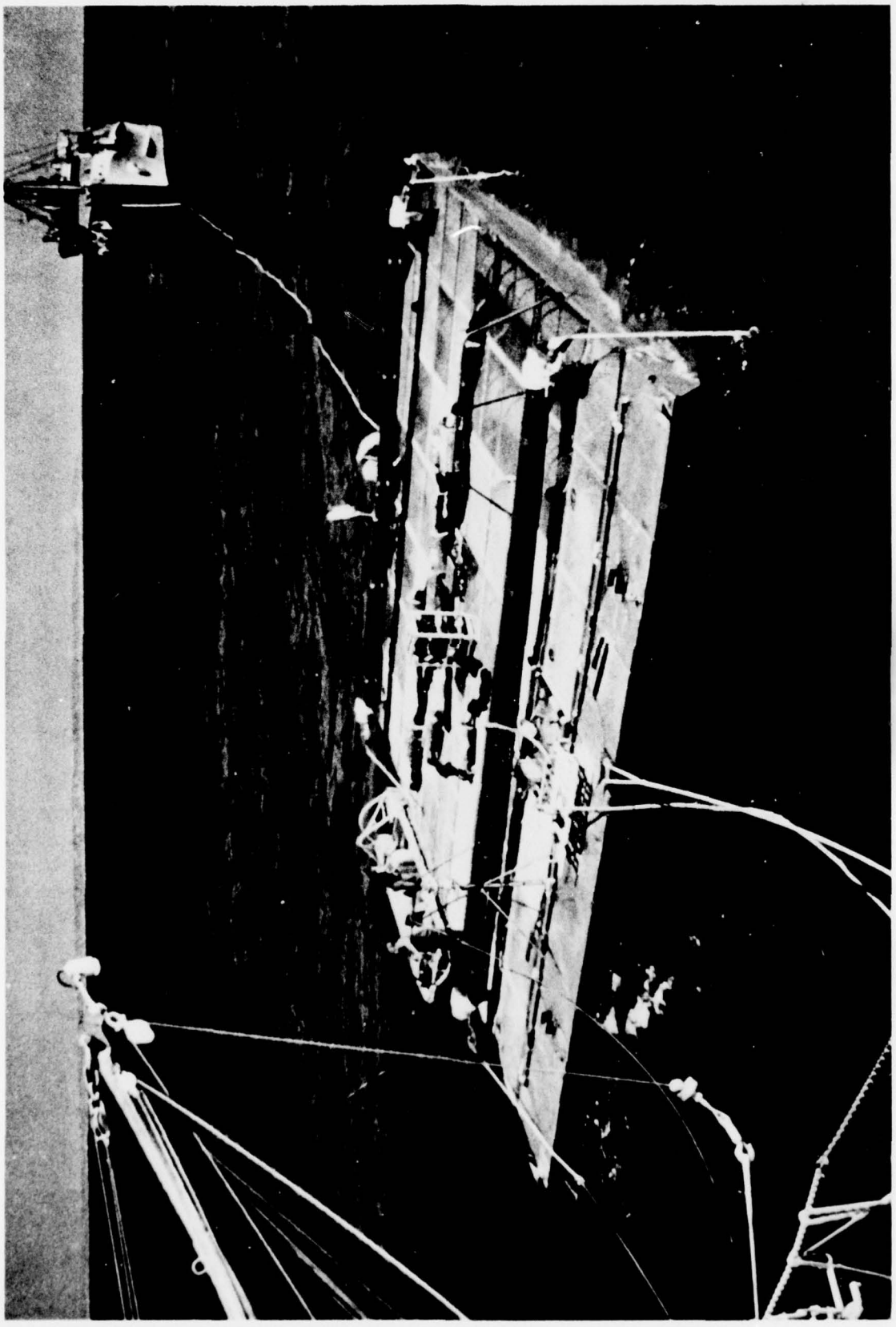


Figure 4. Barge at sea

LIST OF ATTENDEES

* Author

Mr. L.L. Anderson	W.R.E. Salisbury
Mr. J. W. Archer	Sydney University
Mr. R.W. Arstall	W.R.E. Salisbury
* Mr. M.W. Asten	Macquarie University
Dr. P.W. Baker	W.R.E. Salisbury
Mr. J.S. Bell	W.R.E. Salisbury
* Professor T.G. Birdsall	University of Michigan, Michigan U.S.A.
Mr. V.J. Bosher	W.R.E. Salisbury
Dr. I.A. Bourne	University of Melbourne
Mr. K. Breynard	W.R.E. Salisbury
Mr. M.S. Brown	W.R.E. Salisbury
Dr. A.S. Burgess	W.R.E. Salisbury
* Mr. J.M. Burton	Telecom Australia, Melbourne
Mr. M.A. Cameron	Australian National University, Canberra
Dr. A. Cantoni	University of Newcastle
Dr. W. Carroll	Western Australian Institute Technology
Dr. D.G. Cartwright	W.R.E. Salisbury
Dr. I. Chessell	W.R.E. Salisbury
Mr. A.M. Collins	Australian National University, Canberra
Dr. C.L. Cook	W.R.E. Salisbury
Mr. R.P. Coutts	University of Adelaide
* Dr. D.F. Crawford	University of Sydney
* Mr. H.A. d'Assumpcao	W.R.E. Salisbury
Mr. R.N. Davie	W.R.E. Salisbury
Dr. B.R. Davis	University of Adelaide
Mr. B.S. Deegan	W.R.E. Salisbury
Mr. F.A. Dixon	W.R.E. Salisbury
* Mr. P.C. Drewer	W.R.E. Salisbury
Mr. L.J. Dunne	W.R.E. Salisbury
Mr. F.B. Dwyer	C.S.R. Research Laboratory, Roseville, N.S.W.
Dr. A.G. Fabula	Naval Undersea Center, San Diego, California U.S.A.
Dr. A.J. Fenwick	R.A.N. Research Laboratory, Edgecliff, N.S.W.
Mr. W.O. Gibberd	W.R.E. Salisbury

LIST OF ATTENDEES (CONTD.)

Dr. G.L. Goodwin	S.A. Institute of Technology
* Dr. D.A. Gray	W.R.E. Salisbury
Mr. G.R. Haack	W.R.E. Salisbury
Mr. I.A. Hagan	R.A.N. Research Laboratory, Sydney, N.S.W.
* Mr. D.N. Hart	Adelaide University
Mr. M. Heigl	W.R.E. Salisbury
Dr. D.J. Heilbronn	W.R.E. Salisbury
Mr. K.C. Hendy	Aeronautical Research Laboratories, Melbourne, VICTORIA
* Mr. J.D. Henstridge	Australian National University, Canberra
Mr. M.W. Higgins	W.R.E. Salisbury
Dr. R.K. Jarrott	W.R.E. Salisbury
Mr. N.K. Jones	W.R.E. Salisbury
* Professor R.G. Keats	University of Newcastle
Professor C.S.L. Keay	University of Newcastle
Mr. M.M. Komesaroff	C.S.I.R.O. Epping, N.S.W.
Mr. N.R. Labrum	C.S.I.R.O. Epping, N.S.W.
Mr. D. Lamb	W.R.E. Salisbury
Mr. J.E.H. Lamprey	W.R.E. Salisbury
Mr. T.A. Leedham	Footscray Institute of Technology, Footscray, VICTORIA
Dr. M.L. Lees	W.R.E. Salisbury
Mr. B.A. McDonald	W.R.E. Salisbury
Dr. N.A. McDonald	Department of Comm. Eng. R.M.I.T. Melbourne, VICTORIA
Mr. R.L. McGowan	W.R.E. Salisbury
Dr. D.J. McLean	C.S.I.R.O. Epping, N.S.W.
Mr. A.R. Mahoney	W.R.E. Salisbury
Dr. J.S. Mainstone	University of Queensland
Mr. G.C. Mountford	W.R.E. Salisbury
Mr. M.J. Miller	Institute of Technology, Adelaide
* Dr. K.J. Muirhead	Research School of Earth Sciences, A.N.U. Canberra
* Mr. J.V. Murphy	Telecom Australia, Melbourne
* Professor G.H. Newstead	Australian National University, Canberra
Dr. D.G. Nichol	W.R.E. Salisbury
Dr. K.B. Parcell	W.R.E. Salisbury

LIST OF ATTENDEES (CONTD.)

Mr. D.J. Parker	W.R.E. Salisbury
Mr. R. Paterson	W.R.E. Salisbury
Dr. D.J. Patterson	W.R.E. Salisbury
Dr. I.C. Potter	W.R.E. Salisbury
Dr. J.O. Penrose	Western Australian Institute of Technology
Mr. G.A.N. Preston	University of Newcastle
* Mr. Ram Datt	Research School of Earth Sciences, Australian National University, Canberra
Mr. R.D. Ramsay	W.R.E. Salisbury
Mr. P.M. Roberts	W.R.E. Salisbury
* Mr. D.E. Robinson	Ultrasonics Institute, Millers Point, N.S.W.
Mr. A. Ross	PO Cybernetics Group, Aero. Research Laboratory, Melbourne.
. Shackleford	Institute of Technology S.A.
.A. Shaw	Footscray Institute of Technology, Footscray, VICTORIA
Mr. P.L. Silver	W.R.E. Salisbury
Mr. D.J. Skellern	Sydney University
Mr. R.J. Skevington	Western Australian Institute of Technology
Mr. L.B. Soden	W.R.E. Salisbury
Mr. A.K. Steele	W.R.E. Salisbury
* Professor P.L. Stocklin	Loughborough University of Technology, U.K.
Dr. G.R. Sutton	R.A.N. Research Laboratory, N.S.W.
Dr. D.R. Sweet	W.R.E. Salisbury
Mr. G. Thé	University of Tasmania
Mr. J.R. Tilbrook	W.R.E. Salisbury
Mr. K.W.J. Todd	W.R.E. Salisbury
Mr. T.N. Tresidder	W.R.E. Salisbury
Mr. R. Trueblood	Naval Undersea Center, San Diego, California U.S.A.
Mr. I.R. Vickers	C.S.R. Research Laboratory, Roseville, N.S.W.
Mr. B.D. Ward	University of Adelaide
Mr. G. Watts	W.R.E. Salisbury
Mr. P.F. Weyers	MAR Inc., Md., U.S.A.
Mr. D.J. Whitford	W.R.E. Salisbury
Mr. K.B. Whiting	W.R.E. Salisbury
Mr. J.L. Whitrow	W.R.E. Salisbury

LIST OF ATTENDEES (CONTD.)

* Dr. J.P. Wild	C.S.I.R.O. Division of Radiophysics, Epping, N.S.W.
Mr. J.G. Williams	W.R.E. Salisbury
Dr. R.J. Wyber	Central Studies Establishment, A.C.T.
Mr. V.K. Yu	University of Newcastle
Mr. G.M. Ziesing	W.R.E. Salisbury

DOCUMENT CONTROL DATA SHEET

Security classification of this page: UNCLASSIFIED

1 DOCUMENT NUMBERS	2 SECURITY CLASSIFICATION
AR Number: AR-000-564	a. Complete Document: Unclassified
Report Number: WRE-MISC-1	b. Title in Isolation: Unclassified
Other Numbers:	c. Summary in Isolation: Unclassified

3 TITLE: PROCEEDINGS OF SYMPOSIUM - SIGNAL PROCESSING FOR ARRAYS

4 PERSONAL AUTHOR(S):

Edited by H.A. d'Assumpcao

5 DOCUMENT DATE: May 1977

6 6.1 TOTAL NUMBER OF PAGES 200

6.2 NUMBER OF REFERENCES:

7 7.1 CORPORATE AUTHOR(S):

Weapons Research Establishment

7.2 DOCUMENT (WING) SERIES AND NUMBER

8 REFERENCE NUMBERS

a. Task:

b. Sponsoring Agency:

9 COST CODE: 312293

10 IMPRINT (Publishing establishment):

Weapons Research Establishment

11 COMPUTER PROGRAM(S) (Title(s) and language(s))

12 RELEASE LIMITATIONS (of the document):

Approved for public release.

12.0	OVERSEAS	NO		P.R.	1	A		B		C		D		E	

Security classification of this page: UNCLASSIFIED

Security classification of this page: UNCLASSIFIED

13 ANNOUNCEMENT LIMITATIONS (of the information on these pages):

No limitation.

14 DESCRIPTORS:

a. EJC Thesaurus Terms

Seismology
Hydrophone arrays
Direction finding
Tropospheric propagation
Microwaves
Holography
Signal processing
Arrays
Ultrasonic frequencies
Diagnosis
Radio astronomy
Spectrum analysis
Entropy
Seismic arrays

b. Non-Thesaurus Terms

Scar
Imaging systems

15 COSATI CODES:

1711
2001
2014
0811

16 LIBRARY LOCATION CODES (for libraries listed in the distribution):

SW

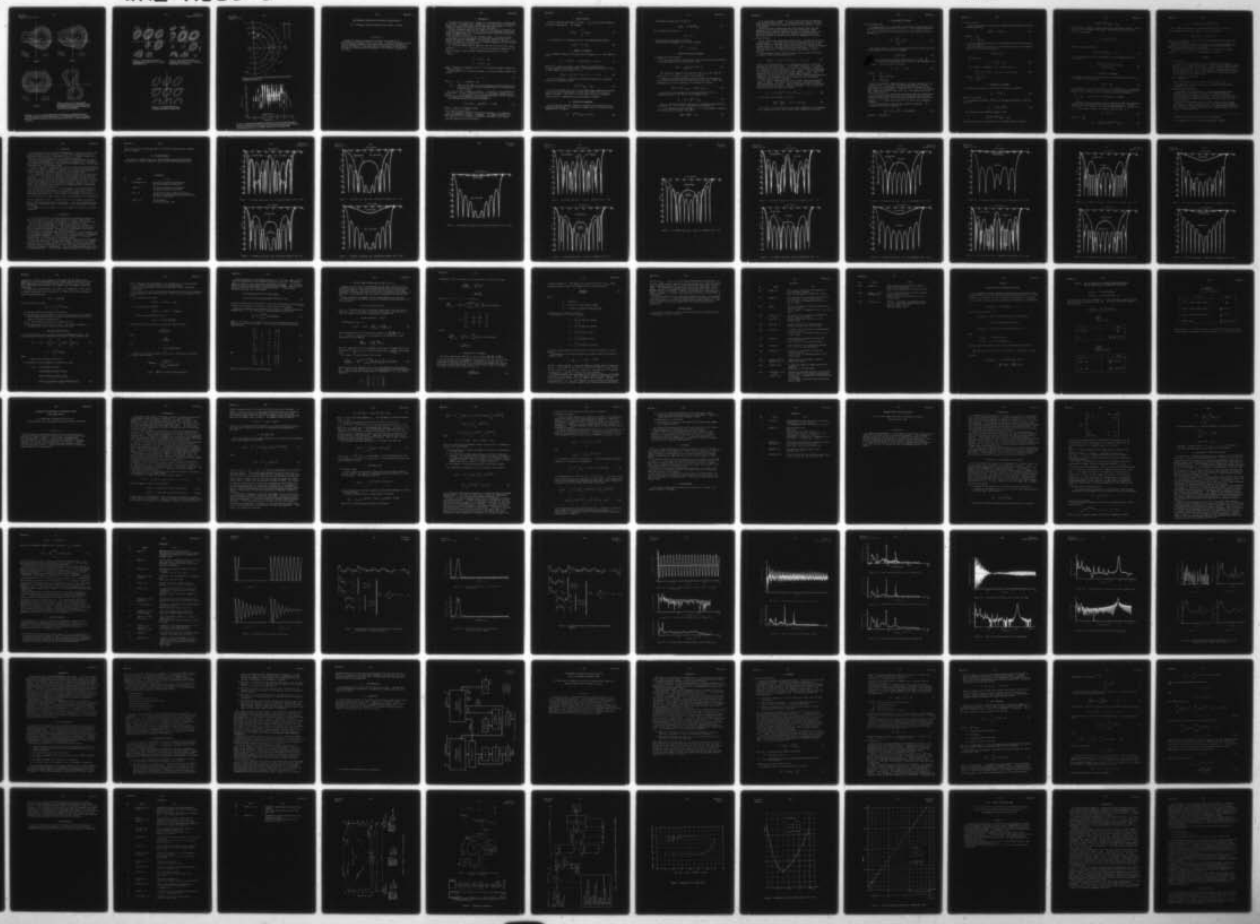
17 SUMMARY OR ABSTRACT:
(if this is security classified, the announcement of this report will be similarly classified)

Security classification of this page: UNCLASSIFIED

AD-A047 017 WEAPONS RESEARCH ESTABLISHMENT SALISBURY (AUSTRALIA) F/G 12/1
PROCEEDINGS OF SYMPOSIUM SIGNAL PROCESSING FOR ARRAYS,(U)
MAY 77 H A D'ASSUMPCAO

UNCLASSIFIED WRE-MISC-1 NL

2 OF 3
ADA
047017



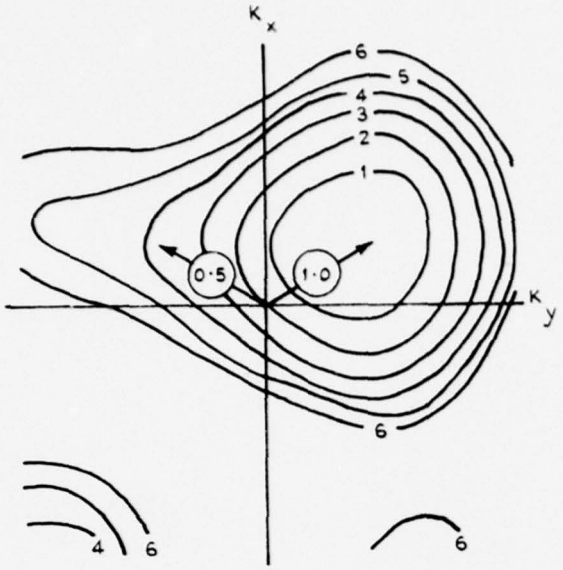


FIGURE 11

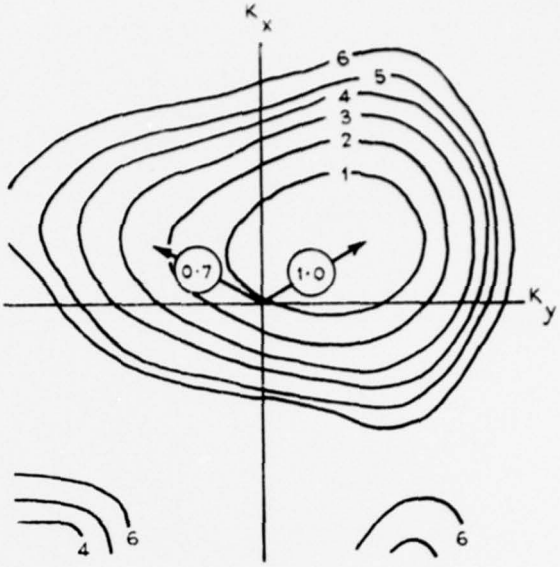


FIGURE 12

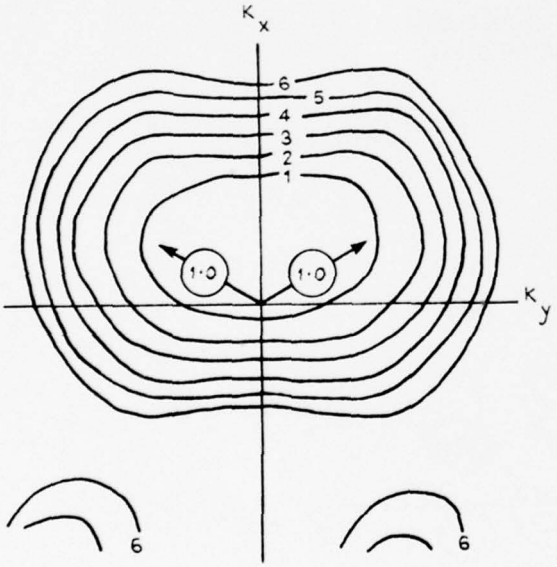


FIGURE 13

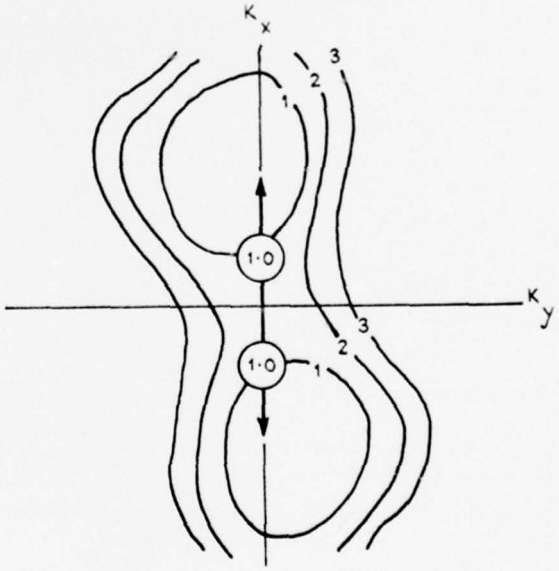


FIGURE 14. TWO 1 Hz, 2 km/s, WAVES, WITH AZIMUTH SEPARATION 180°, INTERFERE TO GIVE WAVENUMBER ESTIMATES BIASED HIGHER THAN THE TRUE VALUES (ARROWED)

FIGURES 11-13. A 1 Hz, 2 km/s WAVE, WITH SOURCE AZIMUTH 60°, IS INTERFERED WITH BY A SECONDARY SOURCE OF RELATIVE AMPLITUDE 0.5, 0.7, 1.0, RESPECTIVELY (The contour maximum shows biased wavenumber estimates, due to incomplete resolution of the 2 sources. Contours are in dB relative to maximum)

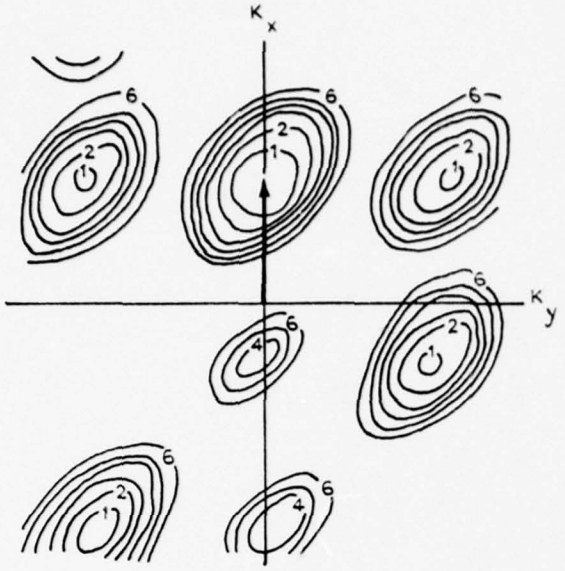


FIGURE 15. WAVENUMBER RESPONSE OF A 4-SEISMOMETER ARRAY TO A 1 Hz, 2 km/s WAVE, WITH SOURCE AT 0°

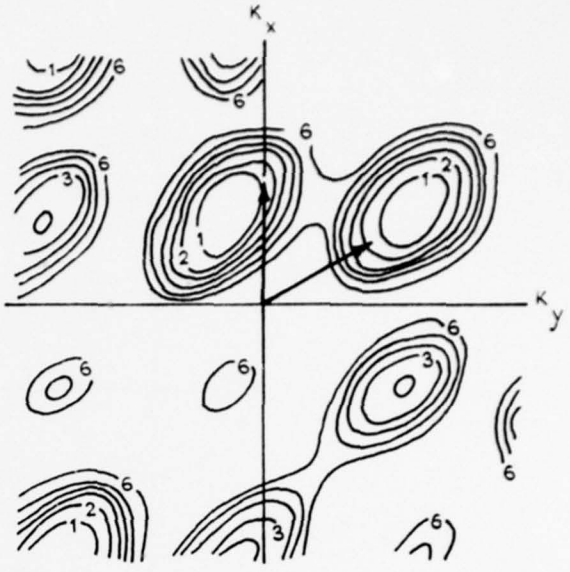


FIGURE 16. WAVENUMBER RESPONSE OF A 4-SEISMOMETER ARRAY TO A 1 Hz, 2 km/s WAVE, WITH SOURCE AT 60° AZIMUTH

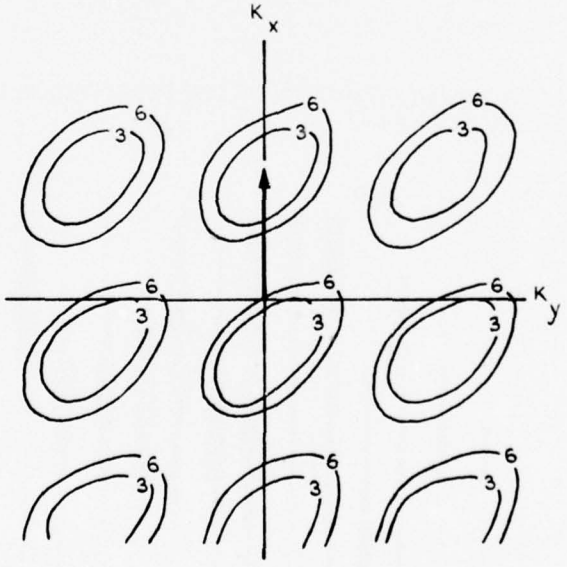


FIGURE 17. WAVENUMBER RESPONSE OF 3-SEISMOMETER ARRAY TO SINGLE WAVE

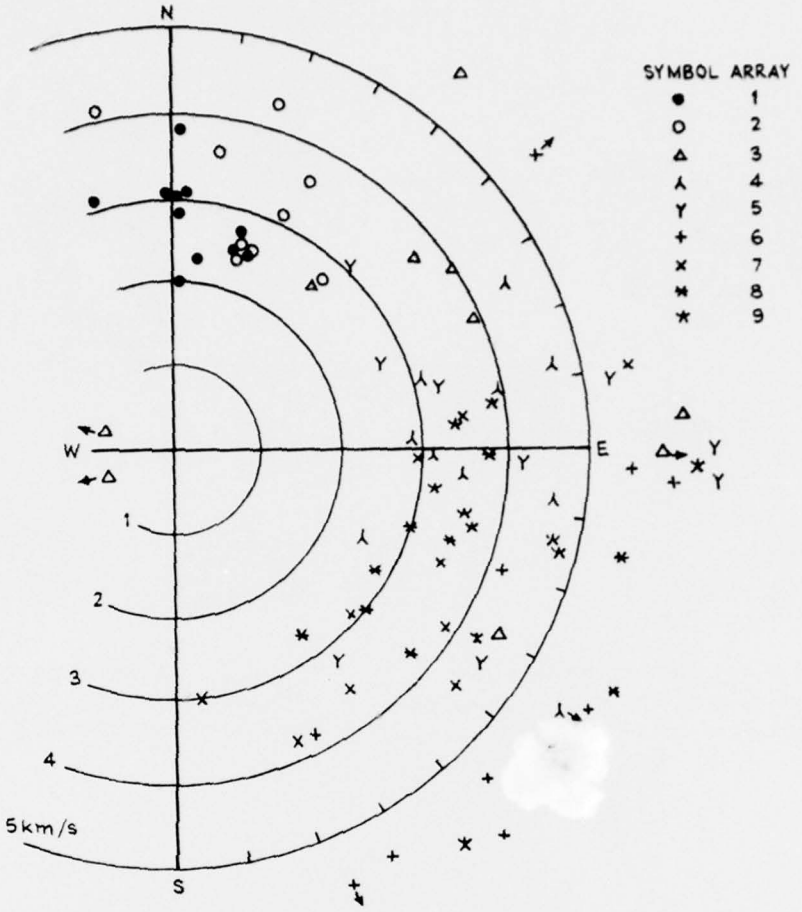


FIGURE 18. VELOCITY-AZIMUTH PLOT FOR ESTIMATES AT 0.49 Hz, WITH 3-SEISMOMETER ARRAYS 1-9

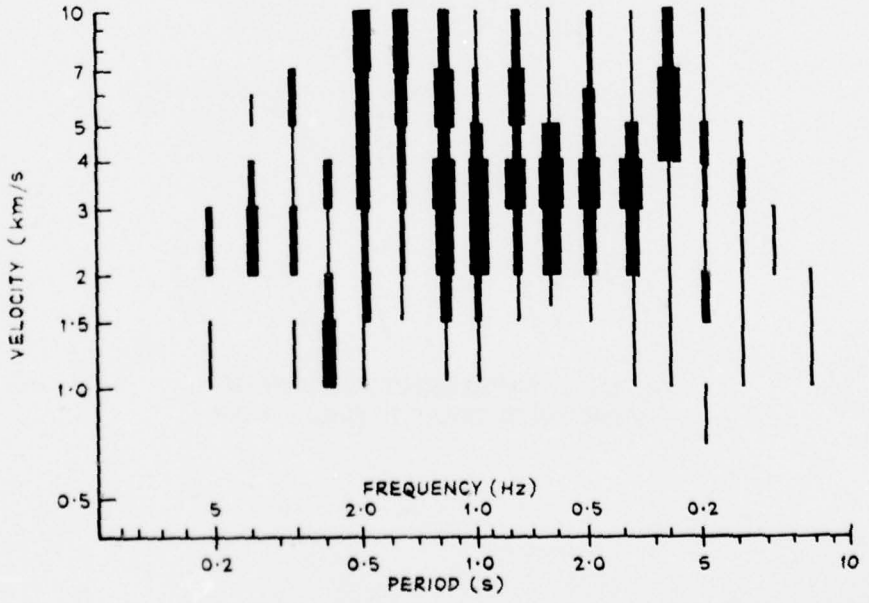


FIGURE 19. HISTOGRAMS OF NUMBER OF ESTIMATES VS VELOCITY, USING 8 VELOCITY INTERVALS ON LOG SCALE (Each histogram refers to estimates in one frequency window. Bar width is approximately proportional to no. of estimates. Maximum no. in each histogram is 80.)

SOME QUADRATIC PROCESSORS FOR ESTIMATING DIRECTIONALITY

H.A. d'Assumpcao, Weapons Research Establishment, Salisbury

S U M M A R Y

Several new quadratic processors are derived for estimating the directionality of signals received by an array. Solutions are given in closed form for Maximum Likelihood, Least-squares, Minimum Bias and Diagonal Fit Quadratic Estimators, as are expressions for the variance of the estimates. The practical advantages and disadvantages of the different techniques are discussed.

1. INTRODUCTION

The directional properties of signals can be measured using a suitable array of receivers. We present here a summary of results for several new processors. Some of the detailed derivations have already been reported(ref.1); it is intended to publish full details of this work separately.

The main assumptions made are that signals are noise-like in character with zero mean and arrive at the array of K receivers as N independent plane wavefronts from known directions. We confine our attention to cases in which the number of signal sources $N \geq K$.

The analysis will be done in the frequency domain. For many practical applications, the spectral components at different frequencies are, to all intents and purposes, independent. We assume this to be the case here. This enables the analysis to be carried out at a single frequency, with the extension to all frequencies then being trivial.

Let x_j be the complex spectral component from the j-th receiver at a given frequency, $j = 1, \ldots, K$. Similarly, let y_k be the spectral component of the signal received at some arbitrary reference point from the k-th source, $k = 1, \ldots, N$.

Write

$$\underset{\sim}{y}^T = \{y_1, y_2, \ldots y_N\},$$

$$\underset{\sim}{x}^T = \{x_1, x_2, \ldots x_K\},$$

where $.^T$ denotes the transpose, and $.^H \equiv .^{*T}$ the complex conjugate transpose of a matrix or vector.

As the signals arrive as plane wavefronts, $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are linearly related, and we have

$$\underset{\sim}{x} = \underset{\sim}{V}\,\underset{\sim}{y}, \tag{1}$$

where $\underset{\sim}{V} = [v_{jk}] = [\exp(i\phi_{jk})]$, and

ϕ_{jk} = phase of the signal (at a given frequency) received from the k-th source by the j-th receiver, measured relative to some arbitrary reference point.

$\underset{\sim}{V}$ is (K x N) and is assumed to be known. In practice, it will generally be of full rank. If not, it implies the array has been badly designed, for if rank $(\underset{\sim}{V}) = L < K$, then only L entries in $\underset{\sim}{x}$ are independent, and the remaining (K - L) entries can be deduced without measurement.

The covariance matrix of the receiver outputs is

$$\underset{\sim}{M} \triangleq \langle \underset{\sim}{x}\,\underset{\sim}{x}^H \rangle = \underset{\sim}{V} \langle \underset{\sim}{y}\,\underset{\sim}{y}^H \rangle \underset{\sim}{V}^H = \underset{\sim}{V}\,\underset{\sim}{S}\,\underset{\sim}{V}^H, \tag{2}$$

where $\langle . \rangle$ denotes an ensemble average,

and we have written $\langle \underset{\sim}{y}\,\underset{\sim}{y}^H \rangle \triangleq \underset{\sim}{S}$.

$\underset{\sim}{S}$ is the covariance matrix of the plane-wave arrivals. By assumption, these are independent, so that $\underset{\sim}{S}$ is diagonal. The diagonal elements of $\underset{\sim}{S}$, $\{s_n\}$, which we wish to estimate, correspond to the powers of the signals arriving from each of the N directions.

2. MATRIX NOTATION

We shall adopt the convention of writing $(\cdot)_{mn}$ for the (m,n)th element of the matrix in parentheses. For example,

$$(\underset{\sim}{A}\ \underset{\sim}{B})_{ij} = \sum_{k} a_{ik}\, b_{kj}. \tag{3}$$

If $\underset{\sim}{A}$ and $\underset{\sim}{B}$ are (m x n) matrices, we write the Hadamard direct product

$$\underset{\sim}{A} \boxdot \underset{\sim}{B} = [a_{ij}\, b_{ij}]. \tag{4}$$

3. QUADRATIC ESTIMATORS

The estimators examined in this paper are all quadratic, i.e., they take the form

$$\hat{s}_n = \underset{\sim}{x}^H \underset{\sim}{A}^{(n)} \underset{\sim}{x} = \mathrm{Tr}(\underset{\sim}{A}^{(n)} \underset{\sim}{x}\ \underset{\sim}{x}^H),\ 1 \leq n \leq N, \tag{5}$$

where $\underset{\sim}{A}^{(n)}$ is some K x K matrix, and $\hat{}$ denotes an estimated value.

The estimator is said to be unbiased if the expected value of each of the N estimates equals the true value; i.e., if

$$\langle \hat{s}_n \rangle = s_n\ \forall\ \{s_n\},\ 0 \leq s_n < \infty,\ n = 1,\ldots,N. \tag{6}$$

It can be shown that a necessary and sufficient condition for a quadratic estimator to be unbiased is that

$$(\underset{\sim}{V}^H \underset{\sim}{A}^{(n)} \underset{\sim}{V})_{mm} = \delta_{nm}. \tag{7}$$

It can also be shown that, for $N > K$ and under general conditions occurring in practice, no unbiased quadratic estimator exists which only gives non-negative estimates. Only for $N = K$ can we find, for each s_n, a quadratic estimator which is unbiased and yields non-negative estimates.

4. CONVENTIONAL BEAMFORMER

The conventional beamformer, variously called the "time delay and sum" or "unweighted add-squarer", is a quadratic estimator which is in general biased. Mathematically, we compute

$$\hat{s}_n = (\underset{\sim}{V}^H \underset{\sim}{x}\ \underset{\sim}{x}^H \underset{\sim}{V})_{nn},\ 1 \leq n \leq N. \tag{8}$$

The ensemble average of the estimate is

$$\langle \hat{s}_n \rangle = (\underset{\sim}{V}^H \underset{\sim}{\underset{\sim}{V}} \underset{\sim}{S} \underset{\sim}{V}^H \underset{\sim}{V})_{nn}, \tag{9}$$

and is unbiased if and only if

$$\underset{\sim}{V}^H = \underset{\sim}{V}^{-1},$$

a condition that rarely obtains in practice.

In the notation of equation (5), for this estimator

$$\underset{\sim}{A}^{(n)} = [v_{in}\ v_{jn}^*]. \tag{10}$$

5. MAXIMUM LIKELIHOOD ESTIMATOR

5.1 Maximum likelihood equation

We assume here that the (complex) signal arrivals have normal distributions. Their joint probability density function is then

$$p(\underset{\sim}{x}) = \frac{1}{\pi^K |\underset{\sim}{M}|} \exp(-\underset{\sim}{x}^H \underset{\sim}{M}^{-1} \underset{\sim}{x}). \tag{11}$$

(The additional assumptions have been made that $\underset{\sim}{V}$ is of full rank and that $s_n > 0 \ \forall\ n$, so that $\underset{\sim}{M} = \underset{\sim}{V}\ \underset{\sim}{S}\ \underset{\sim}{V}^H$ is non-singular).

To derive the maximum likelihood estimator, we take the logarithm of $p(x)$, differentiate with respect to each s_n and set to zero. After some manipulation, the set of equations to be solved may be written

$$(\underset{\sim}{V}^H \underset{\sim}{M}^{-1}\ \underset{\sim}{x}\ \underset{\sim}{x}^H \underset{\sim}{M}^{-1}\ \underset{\sim}{V})_{nn} = (\underset{\sim}{V}^H \underset{\sim}{M}^{-1}\ \underset{\sim}{V})_{nn} \ \forall\ n. \tag{12}$$

The writer has not succeeded in solving equation (12) for $N > K$. Confining ourselves to the case $N = K$, we readily obtain

$$\hat{s}_n = (\underset{\sim}{V}^{-1}\ \underset{\sim}{x}\ \underset{\sim}{x}^H\ \underset{\sim}{V}^{-1\ H})_{nn}. \tag{13}$$

This is a quadratic estimator, which may be re-arranged to the form given in equation (5). It is evident from equation (13) that $\hat{s}_n \geq 0$.

It is not difficult to show that $\langle \hat{\underset{\sim}{S}} \rangle = \underset{\sim}{S}$; i.e., the estimator is unbiased. The variance of the estimates is given by

$$\langle \hat{s}_n^{\,2} \rangle - \langle \hat{s}_n \rangle^2 = s_n^{\,2}. \tag{14}$$

It is interesting to compare this result with the theoretical Cramer-Rao lower bound on variance(ref.2). It can be shown that, in this case, the actual variance achieves the Cramer-Rao bound, and the estimator is said to be "efficient".

The fact that the standard deviation of the estimate equals the mean demonstrates the intuitively obvious point that in practice a single observation period would be inadequate, and that the variance needs to be reduced by averaging the estimate over a number of observation periods. This general result is true of all the estimators examined in this paper.

5.2 Effect of noise

So far we have been considering an ideal system. In practice, there will be noise present. This could, for example, be amplifier self-noise, or an external signal arriving other than as a plane wave-front from one of the assumed directions. We shall assume that this noise has a normal distribution with zero mean, and is uncorrelated with the other signals. Let its covariance matrix be $\underset{\sim}{N}$. With such extraneous noise present, $\underset{\sim}{x} \neq \underset{\sim}{V}\underset{\sim}{y}$ in general.

If the estimator has been designed without taking account of the noise, the expected value of the estimates will be

$$\left\langle \hat{S}_n \right\rangle = s_n + (\underset{\sim}{V}^{-1}\ \underset{\sim}{N}\ \underset{\sim}{V}^{-1H})_{nn},\ n = 1,2,\ldots,K. \qquad (15)$$

Since $\underset{\sim}{N}$ is a covariance matrix, it is non-negative definite, so that the estimates are biased positively. At the higher frequencies, $\underset{\sim}{V}$ is usually well-conditioned, and the effect of the noise will be minimal. However, as the frequency is reduced, the elements of $\underset{\sim}{V}^{-1}$ increase in absolute magnitude, until ultimately at zero frequency $\underset{\sim}{V}$ becomes singular. Hence at the lower frequencies the effect of the noise becomes magnified greatly. This demonstrates the well-known fact that with any high-resolution array processor operating at low frequencies the uncorrelated noise (such as amplifier self-noise) must be kept to a minimum.

If the covariance matrix $\underset{\sim}{N}$ of the noise is known, steps can be taken to reduce its effect. The writer has not succeeded in solving the maximum likelihood equation with noise. However, from equation (15) it can be seen that the effect of the noise is to bias the estimates. This bias can be subtracted, to give the unbiased estimator:

$$\hat{s}_n = (\underset{\sim}{V}^{-1}\ \underset{\sim}{x}\ \underset{\sim}{x}^{H}\ \underset{\sim}{V}^{-1H} - \underset{\sim}{V}^{-1}\ \underset{\sim}{N}\ \underset{\sim}{V}^{-1H})_{nn},\ n = 1,2,\ldots K. \qquad (16)$$

The variance of the estimates is given by

$$\left\langle \hat{s}_n^{\,2} \right\rangle - \left\langle \hat{s}_n \right\rangle^2 = (S + V^{-1}\ NV^{-1H})_{nn}^{\,2}. \qquad (17)$$

Here again, it can be seen that, at the lower frequencies, as the elements of $\underset{\sim}{V}^{-1}$ assume large values, the variance of the estimates will be large.

6. LEAST-SQUARES ESTIMATOR

6.1 Least-squares fit

The approach here is to start with the observed covariance matrix, $\hat{\underset{\sim}{M}}$, and to seek that $\hat{\underset{\sim}{S}}$ which gives the best fit to it. For a single observation, $\hat{\underset{\sim}{M}} = \underset{\sim}{x}\,\underset{\sim}{x}^H$; for multiple observations, $\hat{\underset{\sim}{M}}$ would be some averaged covariance matrix such as

$$\hat{\underset{\sim}{M}} = \frac{1}{L}\sum_{i=1}^{L} \underset{\sim}{x}_i\,\underset{\sim}{x}_i^H. \tag{18}$$

The problem, then, is to find the diagonal matrix $\hat{\underset{\sim}{S}}$ which gives the least-squares solution to the inconsistent equation

$$\underset{\sim}{V}\,\hat{\underset{\sim}{S}}\,\underset{\sim}{V}^H \simeq \hat{\underset{\sim}{M}}; \tag{19}$$

i.e., we wish to find that (diagonal) $\hat{\underset{\sim}{S}}$ which minimises $\|\underset{\sim}{V}\,\hat{\underset{\sim}{S}}\,\underset{\sim}{V}^H - \hat{\underset{\sim}{M}}\|^2$. This can be solved by differentiating with respect to $\hat{s}_j$ and equating to zero for each j. The result (if $\underset{\sim}{W}^{-1}$ exists) is

$$\hat{\underset{\sim}{s}} = \{\hat{s}_j\} = \underset{\sim}{W}^{-1}\,\hat{\underset{\sim}{s}}', \tag{20}$$

where $\hat{\underset{\sim}{s}}' = \{\hat{s}'_n\}$,

$$\underset{\sim}{W} = (\underset{\sim}{V}^H\,\underset{\sim}{V}) \circledcirc (\underset{\sim}{V}^H\,\underset{\sim}{V})^*,$$

$$\hat{s}'_n = (\underset{\sim}{V}^H\,\hat{\underset{\sim}{M}}\,\underset{\sim}{V})_{nn}.$$

Rao (3) has derived a lemma which gives a sufficient condition for the existence of $\underset{\sim}{W}^{-1}$; this can be re-stated for our purposes as follows. "Let $\underset{\sim}{V}$ be (K x N) of full rank, and M be the greatest integer such that no M columns of $\underset{\sim}{V}$ are dependent. If $M \geqslant N - K + 1$, then $\underset{\sim}{W}$ is non-singular." Rao's lemma only extends to values of $N < 2K$; non-singular matrices $\underset{\sim}{W}$ also exist for $N \geqslant 2K$.

$\hat{\underset{\sim}{s}}'$ is the output of a conventional (unweighted) add-squarer beam-former. The least-squares estimator may therefore be considered as comprising a conventional add-squarer, followed by another processor which corrects the estimate. For a single observation, $\hat{\underset{\sim}{M}} = \underset{\sim}{x}\,\underset{\sim}{x}^H$, and we have a quadratic estimator.

It is not difficult to show that the estimator is unbiased. The covariance matrix of the estimates is

$$\left\langle (\hat{\underset{\sim}{s}} - \underset{\sim}{s})(\hat{\underset{\sim}{s}} - \underset{\sim}{s})^H \right\rangle = \underset{\sim}{W}^{-1}(\underset{\sim}{H} \circledcirc \underset{\sim}{H}^*)\underset{\sim}{W}^{-1}, \tag{21}$$

where $\underset{\sim}{H} = \underset{\sim}{V}^H\,\underset{\sim}{V}\,\underset{\sim}{S}\,\underset{\sim}{V}^H\,\underset{\sim}{V}$.

6.2 Effect of noise

If noise is present, but has not been allowed for, the estimates will have an expected value of

$$\left\langle \hat{\underline{s}} \right\rangle = \underline{s} + \underline{W}^{-1}\, \underline{n}', \tag{22}$$

where $\underline{n}' = \{n'_j\}$,

$$n'_j = (\underline{V}^H \underline{N}\, \underline{V})_{jj}.$$

As in the maximum likelihood case, the estimates are biased, but here the biases can be negative.

If $\underline{N}$ is known, we can allow for it by selecting $\hat{\underline{S}}$ to give a least-squares solution to the inconsistent equation

$$\underline{V}\, \hat{\underline{S}}\, \underline{V}^H + \underline{N} \simeq \hat{\underline{M}}.$$

The solution is

$$\hat{\underline{s}} = \hat{\underline{W}}^{-1} (\hat{\underline{s}}' - \underline{n}'). \tag{23}$$

This estimator is unbiased, with a covariance matrix

$$\left\langle (\hat{\underline{s}} - \underline{s})(\hat{\underline{s}} - \underline{s})^H \right\rangle = \underline{W}^{-1} \{ (\underline{H} + \underline{J}) \boxdot (\underline{H} + \underline{J})^* \} \underline{W}^{-1}, \tag{24}$$

where $\underline{H} = \underline{V}^H \underline{V}\, \underline{S}\, \underline{V}^H \underline{V}$

$\underline{J} = \underline{V}^H \underline{N}\, \underline{V}$.

7. MINIMUM BIAS ESTIMATOR

We confine our attention in this section to quadratic estimators

$$\hat{s}_n = \underline{x}^H \underline{A}^{(n)}\, \underline{x}, \tag{25}$$

where $\underline{A}^{(n)}$ is of rank 1.

For non-negative estimates, $\underline{A}^{(n)}$ must be non-negative definite, so that we can write

$$\hat{s}_n = \underline{x}^H \underline{a}^{(n)}\, \underline{a}^{(n)H}\, \underline{x}. \tag{26}$$

The condition for an unbiased estimator is that

$$(\underline{V}^H \underline{a}^{(n)}\, \underline{a}^{(n)H}\, \underline{V})_{mm} = \delta_{nm},$$

but we know (see Section 3) that for $N > K$ it cannot be met, so we seek

$$(\underset{\sim}{V}^H \underline{a}^{(n)} \underline{a}^{(n)H} \underset{\sim}{V})_{mm} \simeq \delta_{nm}.$$

We derive here a quadratic estimator which gives non-negative estimates, and for which the bias is minimum in some sense. Specifically, we wish to find $\underline{a}^{(n)}$ so as to minimise

$$\sum_n \mathrm{Tr}(\underset{\sim}{V}^H \underline{a}^{(n)} \underline{a}^{(n)H} \underset{\sim}{V})$$

subject to the constraint that

$$(\underset{\sim}{V}^H \underline{a}^{(n)} \underline{a}^{(n)H} \underset{\sim}{V})_{nn} = 1 \ \forall \ n. \tag{27}$$

This can be solved in a straightforward way using Lagrange multipliers. The solution is

$$\hat{s}_n = \frac{\{\underset{\sim}{V}^H(\underset{\sim}{V}\ \underset{\sim}{V}^H)^{-1}\ \underline{x}\ \underline{x}^H(\underset{\sim}{V}\ \underset{\sim}{V}^H)^{-1}\ \underset{\sim}{V}\}_{nn}}{\{\underset{\sim}{V}^H(\underset{\sim}{V}\ \underset{\sim}{V}^H)^{-1}\underset{\sim}{V}\}^2_{nn}} . \tag{28}$$

8. DIAGONAL FIT ESTIMATOR

Consider an observed or estimated covariance matrix $\hat{\underset{\sim}{M}}$, and the case $N > K$. We can find a matrix $\hat{\underset{\sim}{S}}$ such that

$$\underset{\sim}{V}\ \hat{\underset{\sim}{S}}\ \underset{\sim}{V}^H \equiv \hat{\underset{\sim}{M}}, \tag{29}$$

but in general only if $\hat{\underset{\sim}{S}}$ is not diagonal. The approach here is to allow $\hat{\underset{\sim}{S}}$ to have non-zero off-diagonal elements so that we can obtain a general solution to equation (29); we then select the particular solution which minimises

$\epsilon \triangleq \sum_{i \neq j} |\hat{s}_{ij}|^2$. We call this the "Diagonal Fit" Estimator.

The general solution to equation (29) is expressed in terms of a generalised inverse of $\underset{\sim}{V}$ and an arbitrary complex matrix $\underset{\sim}{Z}$(ref.4,page 30). We then differentiate ϵ with respect to the real and imaginary parts of z_{ij} and set to zero. The eventual solution, if the matrix inverses exist, is

$$\hat{\underline{s}} = \underset{\sim}{Q}^{-1}\ \hat{\underline{d}} \tag{30}$$

where $\hat{\underline{d}} = \{\hat{d}_n\}$,

$$\hat{d}_n = \{\underset{\sim}{V}^H(\underset{\sim}{V}\ \underset{\sim}{V}^H)^{-1}\ \hat{\underset{\sim}{M}}(\underset{\sim}{V}\ \underset{\sim}{V}^H)^{-1}\ \underset{\sim}{V}\}_{nn}, \tag{31}$$

$$\underset{\sim}{Q} = \{\underset{\sim}{V}^H(\underset{\sim}{V}\ \underset{\sim}{V}^H)^{-1}\underset{\sim}{V}\} \boxdot \{\underset{\sim}{V}^H(\underset{\sim}{V}\ \underset{\sim}{V}^H)^{-1}\underset{\sim}{V}\}^*.$$

Rao's lemma (see Section 6.1) can be applied to $\underset{\sim}{Q}$ as well: since $\underset{\sim}{V}$ is (by assumption) of full rank, $(\underset{\sim}{V}\ \underset{\sim}{V}^H)$ is p.d., so we can write

$$\underset{\sim}{V}^H(\underset{\sim}{V}\ \underset{\sim}{V}^H)^{-1}\underset{\sim}{V} = (\underset{\sim}{F}\ \underset{\sim}{V})^H\ (\underset{\sim}{F}\ \underset{\sim}{V}),$$

where $\underset{\sim}{F}$ is non-singular. Hence if $\underset{\sim}{V}$ meets the conditions of the lemma, both $\underset{\sim}{W}$ and $\underset{\sim}{Q}$ are non-singular.

Comparing equations (30) and (31) with (28), we find that the Diagonal Fit Estimator may be considered to be the Minimum Bias Estimator followed by a further processor which refines the estimate.

The Diagonal Fit Estimator is unbiased and can in general give negative estimates. For $\hat{\underset{\sim}{M}} = \underset{\sim}{x}\ \underset{\sim}{x}^H$, it is another quadratic estimator.

9. EXAMPLES

9.1 Polar response

All the estimators examined here may be considered to be types of beamformers*. As such, it is possible to plot their polar response or beam pattern, i.e., the output of one beam as a signal source (in the far field) moves around the array. This has been done for a circular ring array of 10 equi-spaced receivers, with a signal source in the same plane as the array. The results are given in figures 1 to 18 for the four types of processors and for different values of R/λ, where

R = radius of the array and
λ = wavelength.

For this array, when the inter-element spacing = $\frac{\lambda}{2}$, $\frac{R}{\lambda} = 0.796$.

For comparison, the polar response of a conventional (add-squarer) beamformer is also given.

9.2 Maximum Likelihood Beamformer

The Maximum Likelihood Estimator is a kind of weighted add-squarer beamformer, where the weights are complex and frequency-dependent. Its beam patterns are shown in figures 1 to 5 for $N = K = 10$, and for a beam pointed radially through one receiver. The plane-wave signal arrivals have been taken to be equally spaced around the array, and to be in the plane of the array.

Note that in each pattern there are 9 nulls, always at the (assumed) directions of the other 9 signal arrivals. This result follows because the estimator is unbiased and applies as well to other unbiased estimators**. The processor may therefore be considered as a beamformer which steers N-1 nulls in selected directions.

---

* Hereafter, the two terms will be used interchangeably.

** I am indebted to Dr. D.A. Gray for providing this insight.

When the inter-element spacing is $\lambda/2$ ($R/\lambda = 0.796$), the Conventional Beamformer has a large back lobe (at 180^{o}) but a somewhat narrower main lobe than the Maximum Likelihood Beamformer. At lower frequencies, however, the latter shows quite distinct improvements over the Conventional processor. Because the nulls are always in the same angular positions, its beam pattern is practically invariant with decreasing frequency, while that of the Conventional Beamformer rapidly deteriorates as frequency is reduced. At low frequencies, the Maximum Likelihood Beamformer is a type of super-directive processor whose pattern is sensibly constant with frequency. Of course, one cannot in practice operate at very low frequencies because the magnitude of the weights becomes very large, and the processor becomes unduly sensitive to noise.

9.3 Least-Squares Beamformer

As mentioned in Section 6, the Least-Squares Beamformer comprises a Conventional Beamformer whose outputs are further processed. Its beam patterns are plotted in figures 6 to 8 for $N = 12$. Here again, the signal arrivals are assumed equally spaced in angle.

As expected, there are $N-1 = 11$ nulls in constant directions, and the main lobe is correspondingly narrow, but the side lobes are of comparable magnitude to those of the Conventional Beamformer.

At low frequencies, the magnitude of the weighting coefficients rises very rapidly, so that this processor would only have very limited application as a superdirective beamformer (at least, for this array configuration).

9.4 Minimum Bias Beamformer

This is another weighted add-squarer. As this estimator is biased, the number and positions of the nulls vary with frequency. Beam patterns are plotted in figures 9 to 13 for $N = 20$ (equally-spaced arrivals). The overall performance of this processor is good. The matrix which requires inverting appears to be better conditioned than the corresponding ones for the other processors.

9.5 Diagonal Fit Beamformer

The Diagonal Fit Beamformer comprises a Minimum Bias Beamformer whose outputs are further processed, to give an unbiased estimator. Its beam patterns are plotted in figures 14 to 18, for $N = 12$. As with the least-squares estimator, there are again $N-1 = 11$ regularly spaced nulls.

For $R/\lambda = 0.796$ (see figure 14), apart from the back lobe, the Diagonal Fit Beamformer has side lobes comparable with those of the Conventional Beamformer. At low frequencies, the improvement is more marked, so this processor could find application as a null-steering superdirective processor. The magnitude of its weighting coefficients does not appear to increase as rapidly as that of the Least Squares Beamformer, as frequency is reduced.

10. SUMMARY OF PROPERTIES

A brief summary of the properties of the estimators is given in Table 1. The Least-Squares estimator can be thought of as comprising the Conventional estimator, which in general gives biased estimates, followed by further processing which makes the estimates unbiased. Similarly, the Diagonal Fit estimator comprises the Minimum Bias estimator followed by further processing.

For $N = K$, the Maximum Likelihood, Minimum Bias and Diagonal Fit estimators are the same.

TABLE 1. SUMMARY OF PROPERTIES

Type	Form	Estimates	Comments
Conventional	$\hat{\underset{\sim}{S}}_{CONV.} = \{[\underset{\sim}{V}^H \underset{\sim}{x} \underset{\sim}{x}^H \underset{\sim}{V}]_{nn}\}$	+	$N \geq K$, biased
Maximum Likelihood	$\hat{\underset{\sim}{S}}_{M.L.} = \{[\underset{\sim}{V}^{-1} \underset{\sim}{x} \underset{\sim}{x}^H \underset{\sim}{V}^{-1H}]_{nn}\}$	+	$N = K$, null steering unbiased
Least Squares	$\hat{\underset{\sim}{S}}_{L.S.} = \underset{\sim}{W}^{-1} \hat{\underset{\sim}{S}}_{CONV.}$	±	$N \geq K$, null steering, unbiased
Minimum bias	$\hat{\underset{\sim}{S}}_{M.B.} = \left\{ \frac{[\underset{\sim}{V}^H(\underset{\sim}{V} \underset{\sim}{V}^H)^{-1} \underset{\sim}{x} \underset{\sim}{x}^H(\underset{\sim}{V} \underset{\sim}{V}^H)^{-1} \underset{\sim}{V}]_{nn}}{[\underset{\sim}{V}^H(\underset{\sim}{V} \underset{\sim}{V}^H)^{-1} \underset{\sim}{V}]^2_{nn}} \right\}$	+	$N \geq K$, biased
Diagonal fit	$\hat{\underset{\sim}{S}}_{D.F.} = \underset{\sim}{Q}^{-1} \underset{\sim}{\Lambda} \hat{\underset{\sim}{S}}_{M.B.}$	±	$N \geq K$, null steering, unbiased

$\underset{\sim}{\Lambda}$ is a diagonal matrix

11. DISCUSSION

The investigation of these Estimators/Beamformers is still at an early stage. Much further work is still required, with other array configurations and in particular with real data in the presence of noise. A detailed comparison is needed with alternative processors, e.g., adaptive processors.

The null-steering properties of the unbiased estimators (Maximum Likelihood, Least-Squares and Diagonal Fit) could find practical application. Although it has no null-steering capability, the Minimum Bias Beamformer appears to have good performance over a wide frequency range.

Of course, the practical implementation of any of these processors would need to involve a computer. Whether any one processor is practicable would depend on many factors, such as the number of receivers, the frequency bandwidth of the signals, and whether the processing needs to be done in real time.

It is important to point out that the matrix $\underline{V}$, on which all the computations are based, is a function only of frequency, array geometry and signal arrival directions. If the latter are known (or can be assumed), then all the matrix calculations can be carried out beforehand and the processor coefficients stored, as they do not depend on the actual signals received at any one time. In this regard, these beamformers differ from other frequently used processors, which require the inversion of some estimate of the covariance matrix, an operation which can have problems in practice.

As pointed out in Section 5.1, it is not adequate to estimate directionality using a single observation vector $\underset{\sim}{x}$ (as the variance of the estimate would be too large), but some average needs to be taken over a number (say, L) of independent observations $\underset{\sim}{x}_i$. If the signals are stationary, this can be done by first calculating the estimated powers, $\{\hat{s}_n\}_i$, corresponding to each observed $\underset{\sim}{x}_i$, and then averaging the $\{\hat{s}_n\}$. Alternatively, one can average the covariance matrix, and only beam-form after L observations. This latter approach has the advantage of requiring less matrix calculations, but at the penalty of needing more storage.

If signals are not stationary, it may be acceptable to average over a band of frequencies.

12. CONCLUSIONS

Of the four processors, the Maximum Likelihood and Minimum Bias appear at this early stage to be the most promising for low-frequency applications. The former would be most useful if interference is present in known directions.

The Least-Squares Beamformer would appear to be useful only over a relatively small frequency band. It has the advantage that it operates on the output of a Conventional Beamformer. One could average the power estimates from a Conventional Beamformer, and then use the Least-Squares processing to generate nulls in the directions of interfering signals.

The Diagonal Fit Beamformer appears to be useful over a broader bandwidth than the Least-Squares Beamformer. The first step in the processing is to implement the Minimum Bias Beamformer, and then to use the Diagonal Fit processing to generate nulls in the directions of interfering signals.

Whether any one of the processors described here is practicable will depend on the particular application - the number of elements in the array, the frequency band, the speed and storage capacity of the computer available and especially on whether results are required in real time. The choice of which one to use naturally must depend on circumstances. The Minimum Bias Beamformer has a good all-round performance. The other three would be attractive in a situation in

which there are many interfering sources in directions which are either known or can be estimated.

13. ACKNOWLEDGEMENTS

The writer is indebted to Dr. D.A. Gray of Weapons Research Establishment and Professor R.G. Keats of Newcastle University for many useful discussions.

REFERENCES

No.	Author	Title
1	d'Assumpcao, H.A.	"Estimation of Sound Directionality". WRE-TN-1369 (WR&D), February 1975
2	Cramer, H.	"Mathematical Methods of Statistics". Princeton University Press, 1946
3	Rao, C.R.	"Estimation of Heteroscedastic Variances". Journal of the American Statistical Association, 65, 161-172, March 1970
4	Searle, S.R.	"Linear Models". John Wiley and Sons, 1971

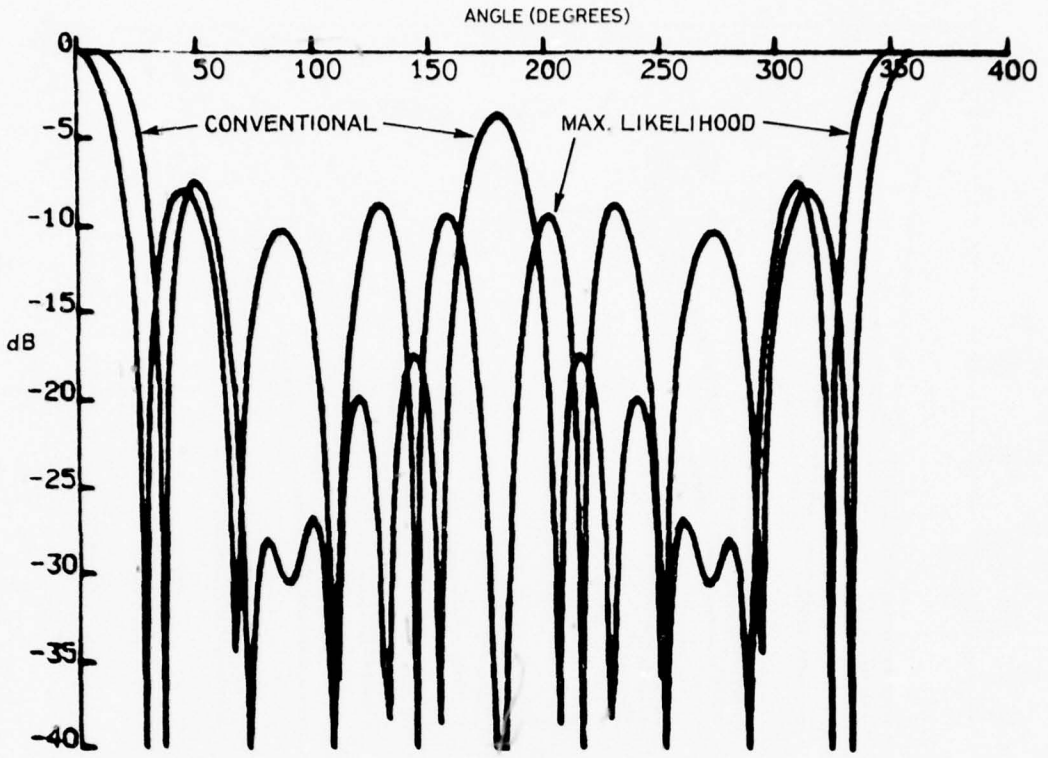


Figure 1. 10-element ring array, max. likelihood estimator, $R/\lambda = 0.796$

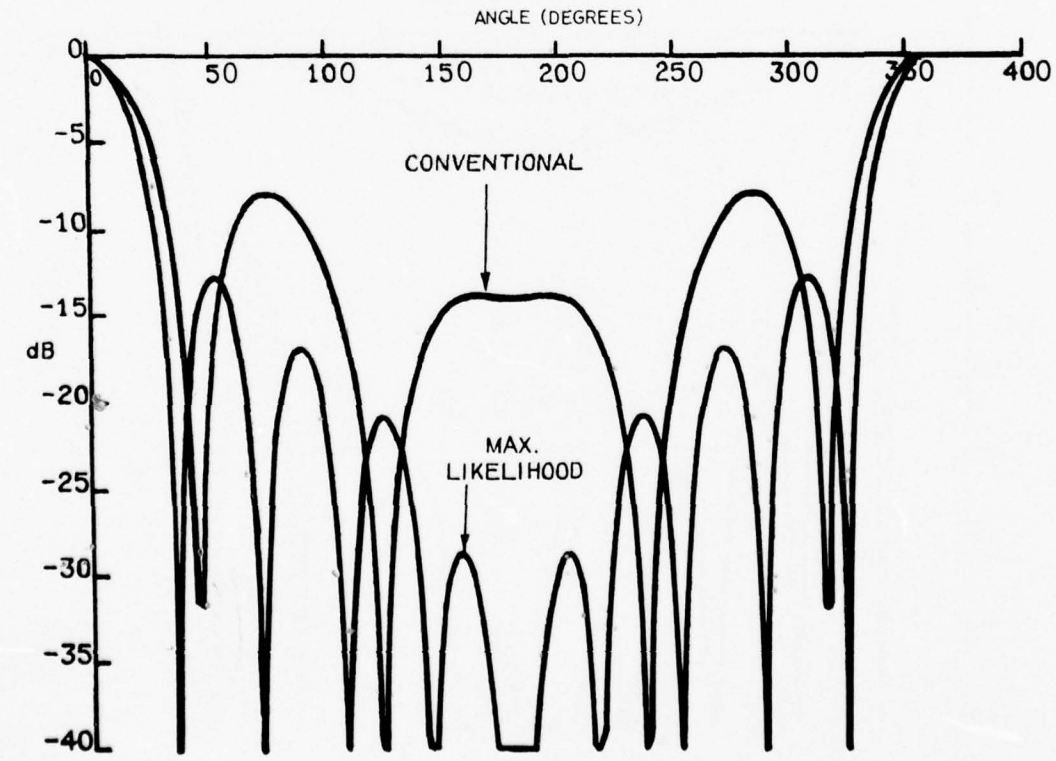


Figure 2. 10-element ring array, max. likelihood estimator, $R/\lambda = 0.5$

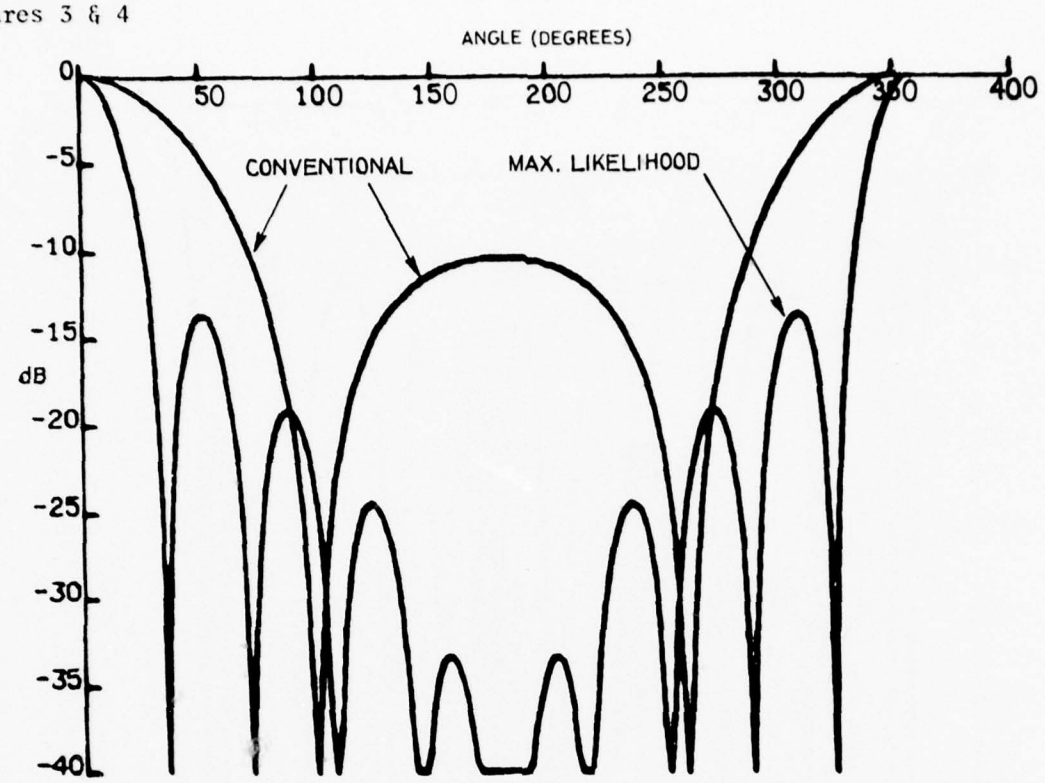


Figure 3. 10-element ring array, max. likelihood estimator, $R/\lambda = 0.25$

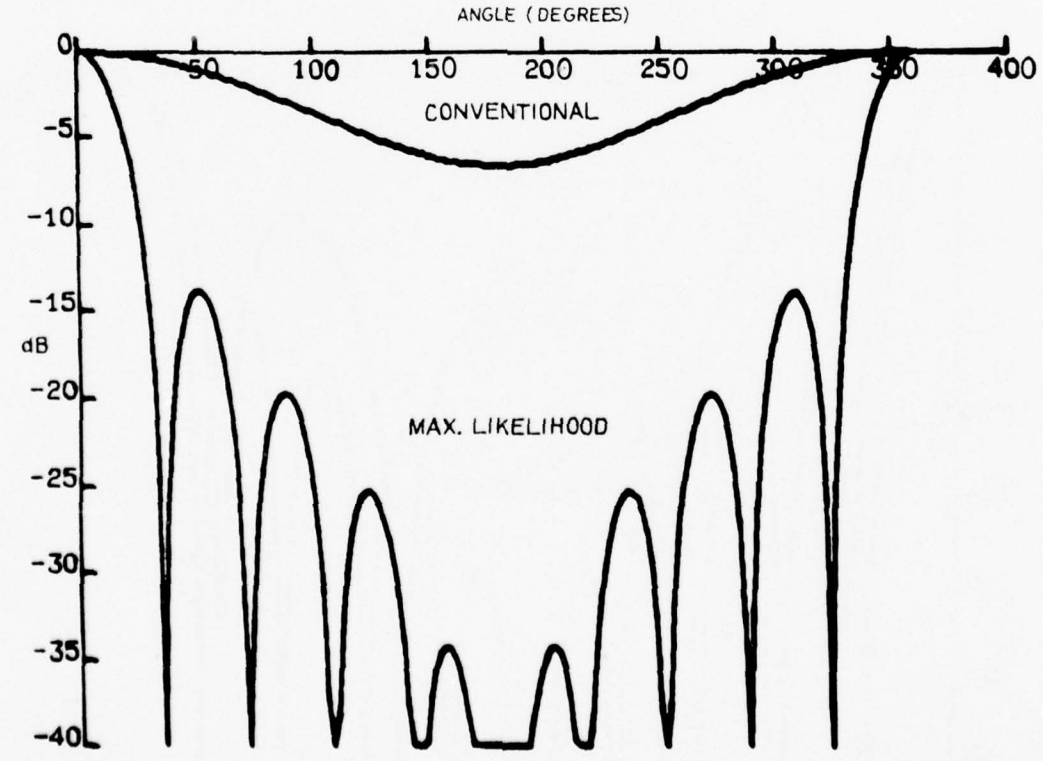


Figure 4. 10-element ring array, max. likelihood estimator, $R/\lambda = 0.125$

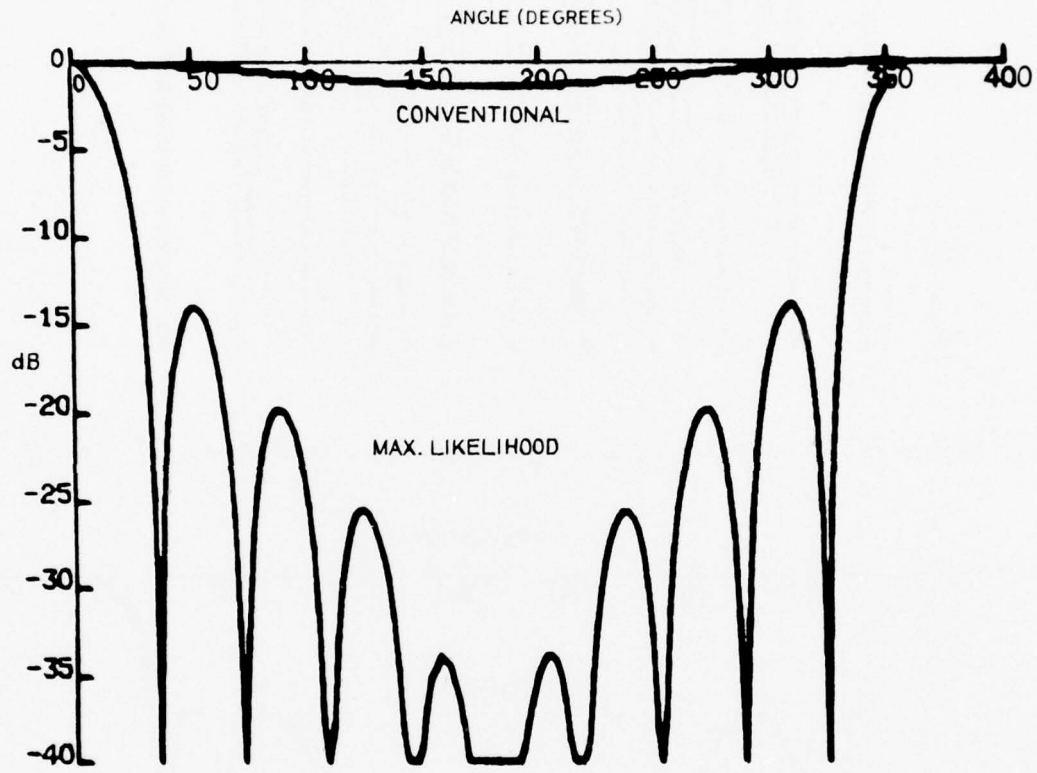


Figure 5. 10-element ring array, max. likelihood estimator, $R/\lambda = 0.0625$

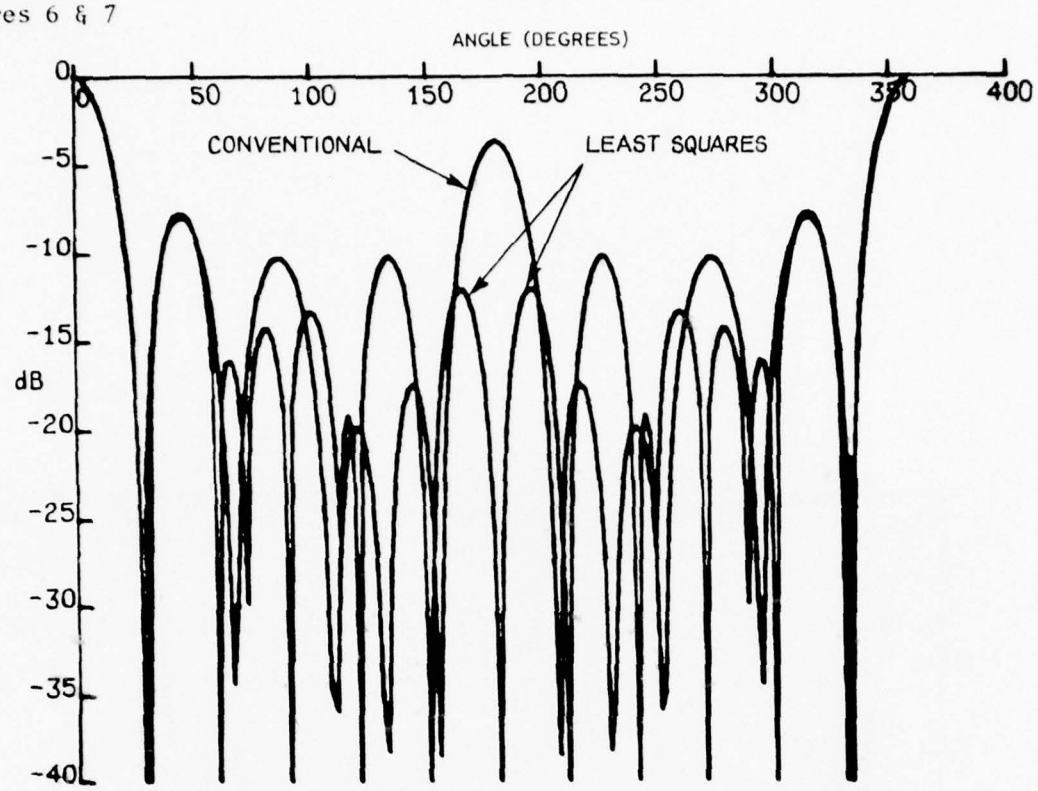


Figure 6. 10-element ring array - least-sq. estimator, $R/\lambda = 0.796$

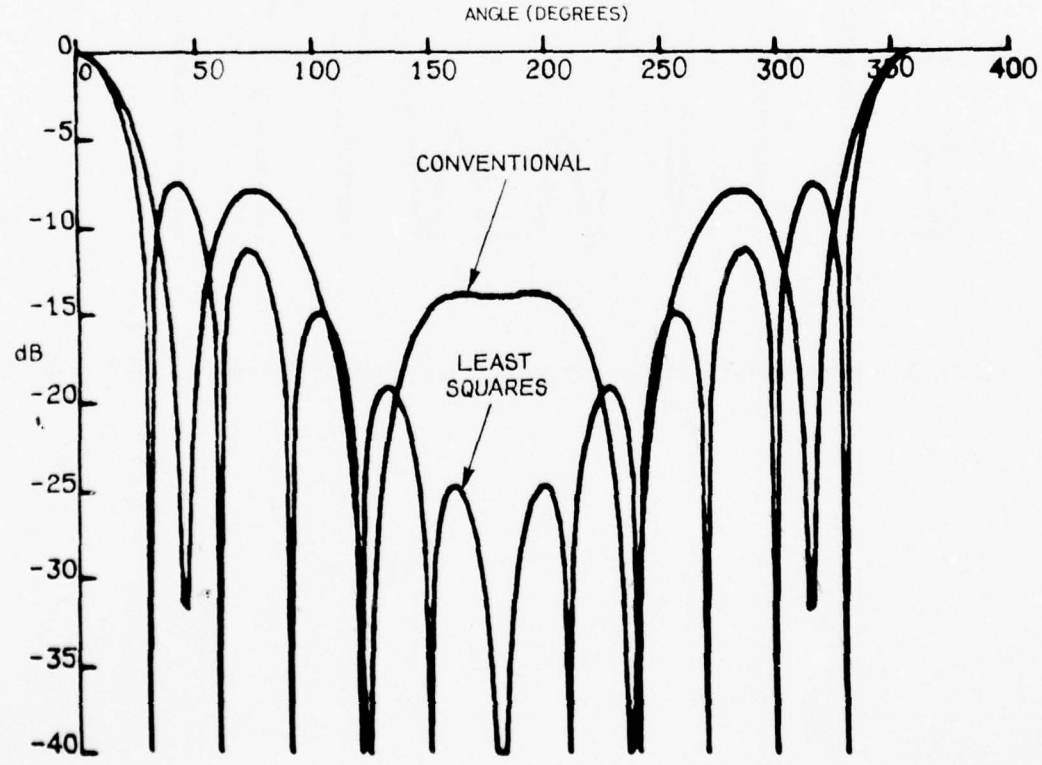


Figure 7. 10-element ring array - least-sq. estimator, $R/\lambda = 0.5$

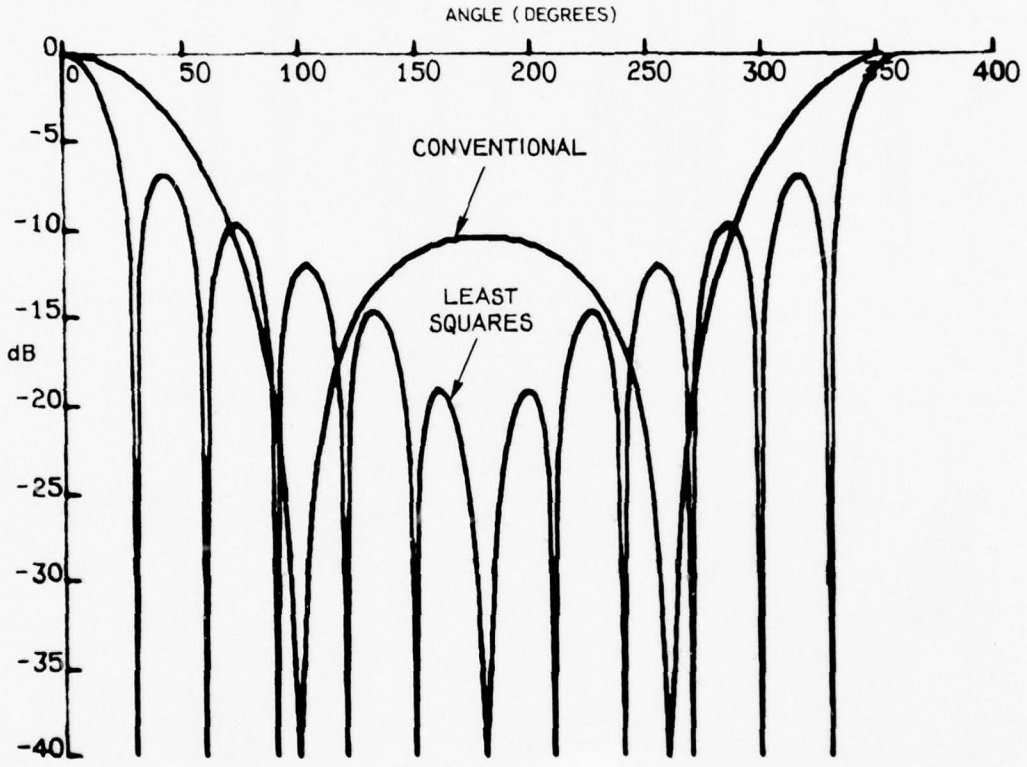


Figure 8. 10-element ring array - least sq. estimator, $R/\lambda = 0.25$

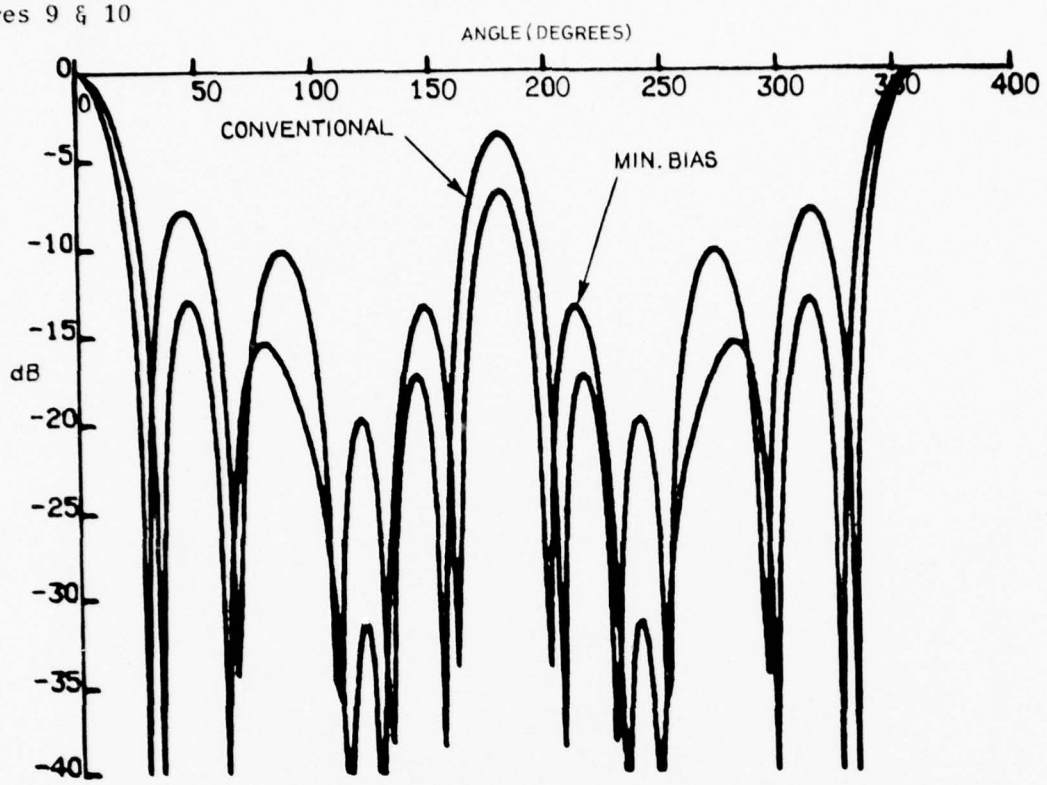


Figure 9. 10-element ring array - min. bias beamformer, $R/\lambda = 0.796$

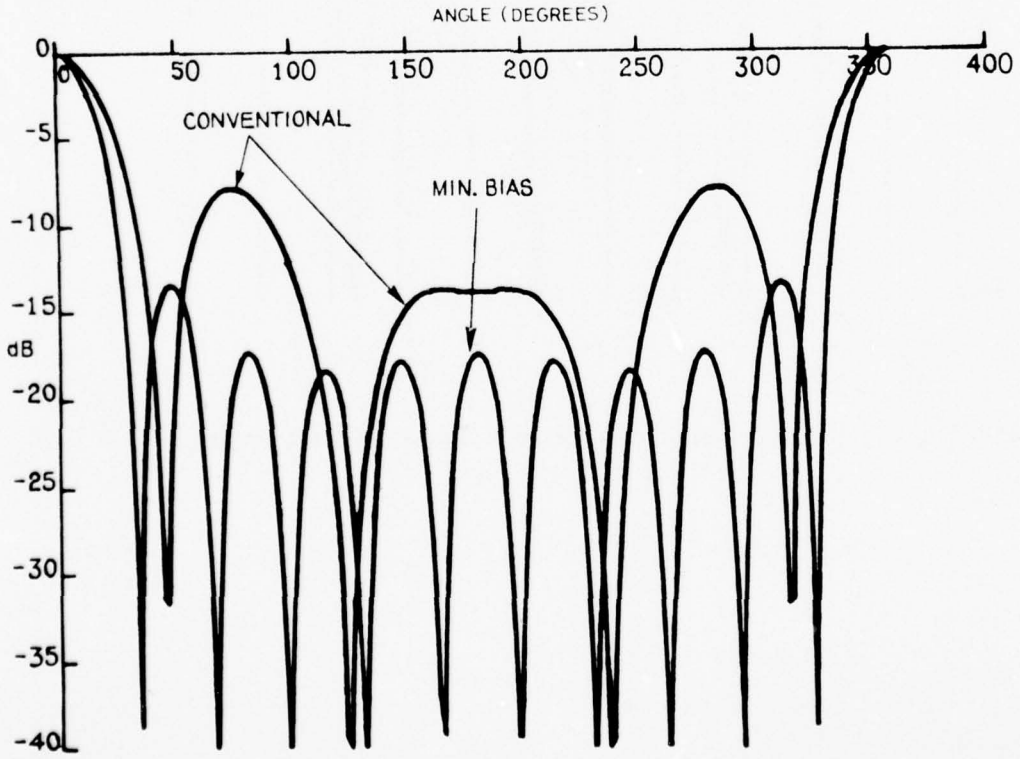


Figure 10. 10-element ring array - min. bias beamformer, $R/\lambda = 0.5$

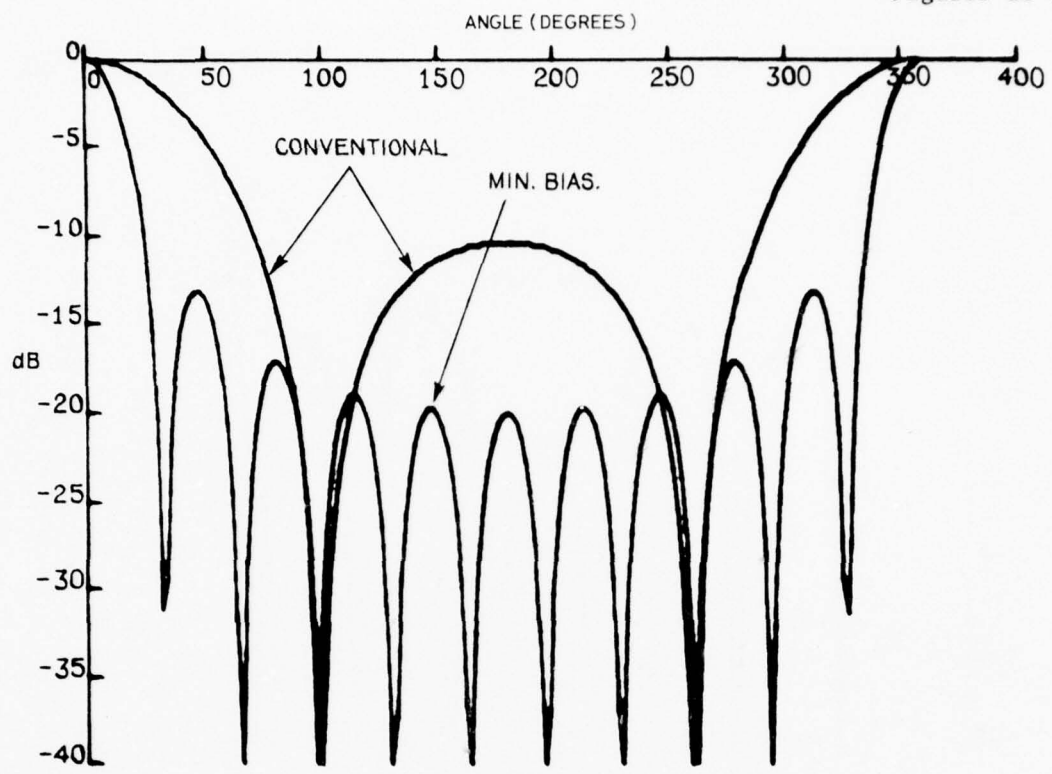


Figure 11. 10-element ring array - min. bias beamformer, $R/\lambda = 0.25$

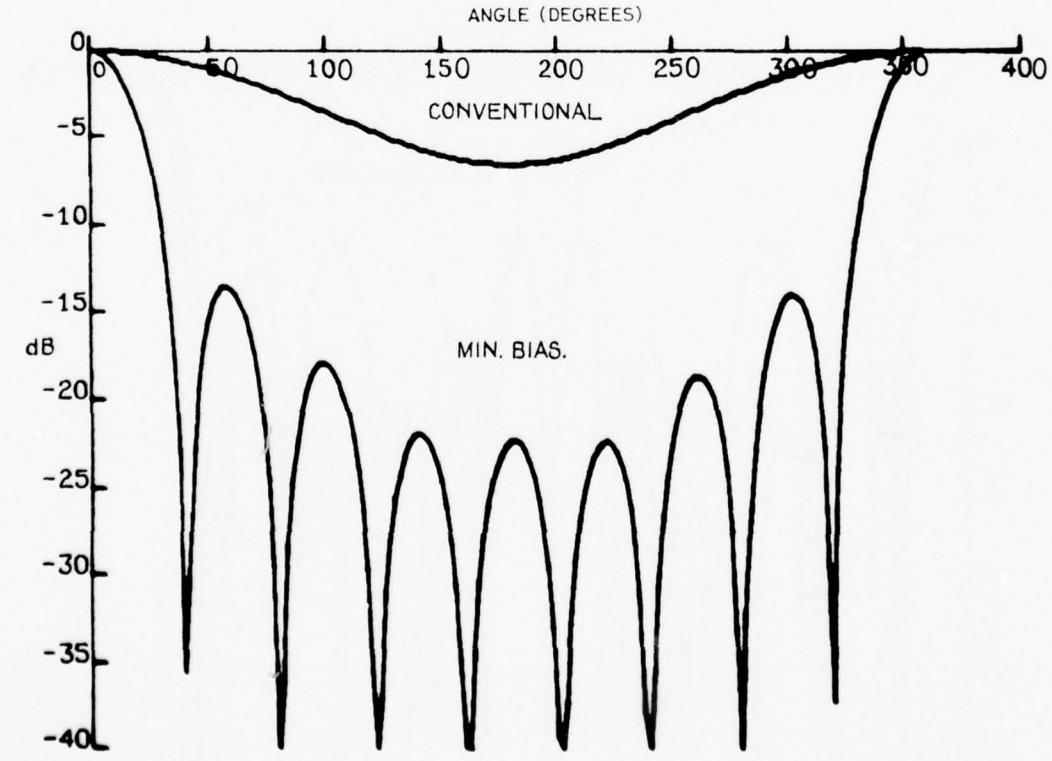


Figure 12. 10-element ring array - min. bias beamformer, $R/\lambda = 0.125$

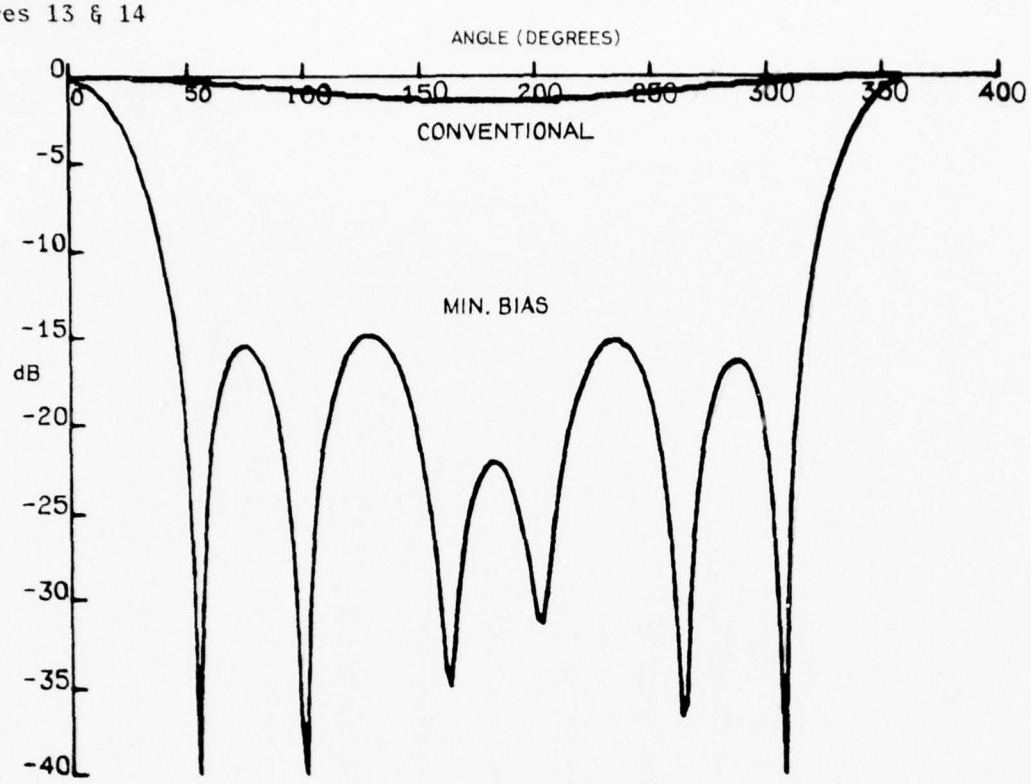


Figure 13. 10-element ring array - min. bias beamformer, $R/\lambda = 0.0625$

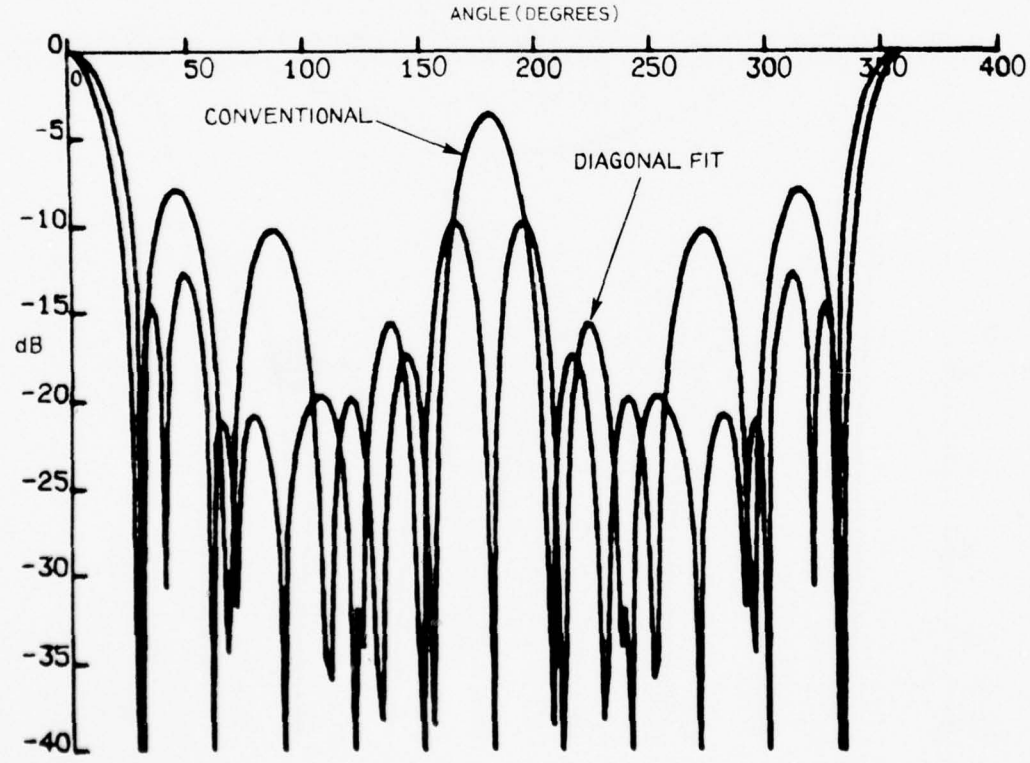


Figure 14. 10-element ring array - diagonal fit beamformer, $R/\lambda = 0.796$

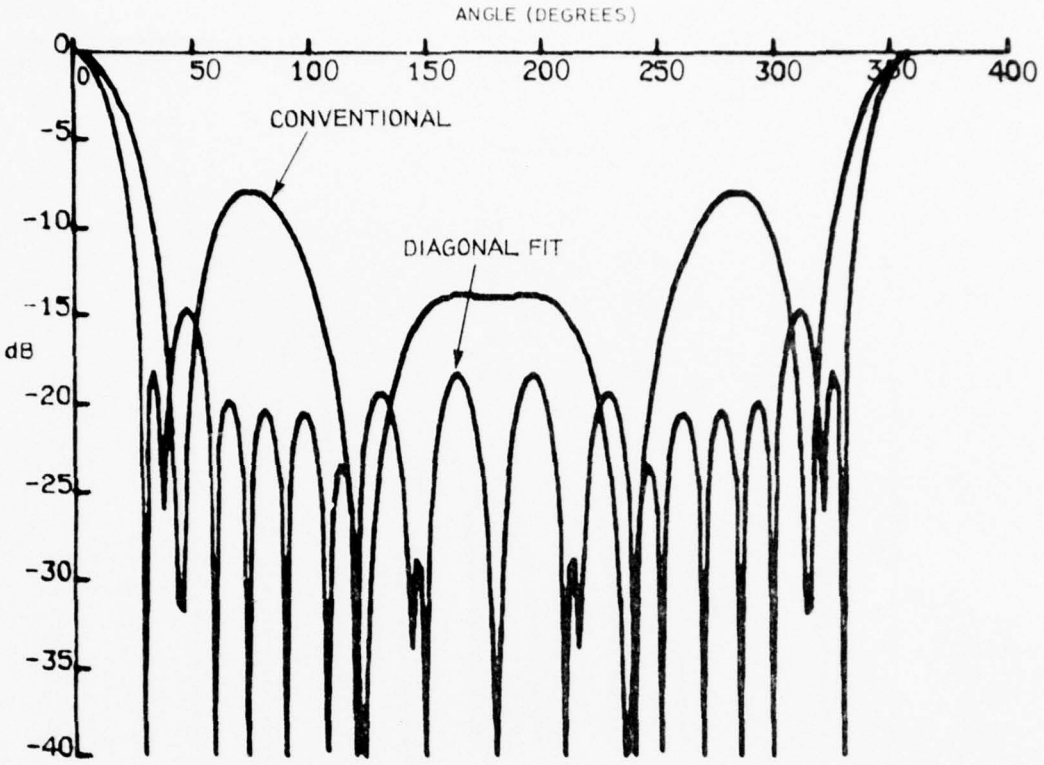


Figure 15. 10-element ring array - diagonal fit beamformer, $R/\lambda = 0.5$

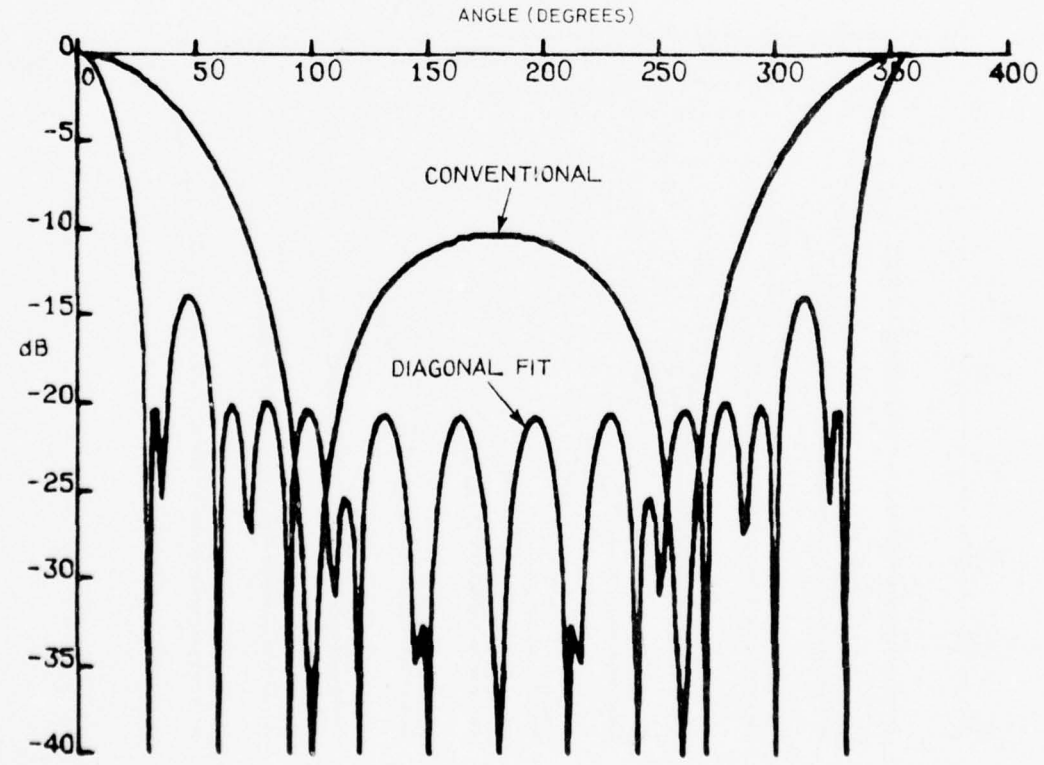


Figure 16. 10-element ring array - diagonal fit beamformer, $R/\lambda = 0.25$

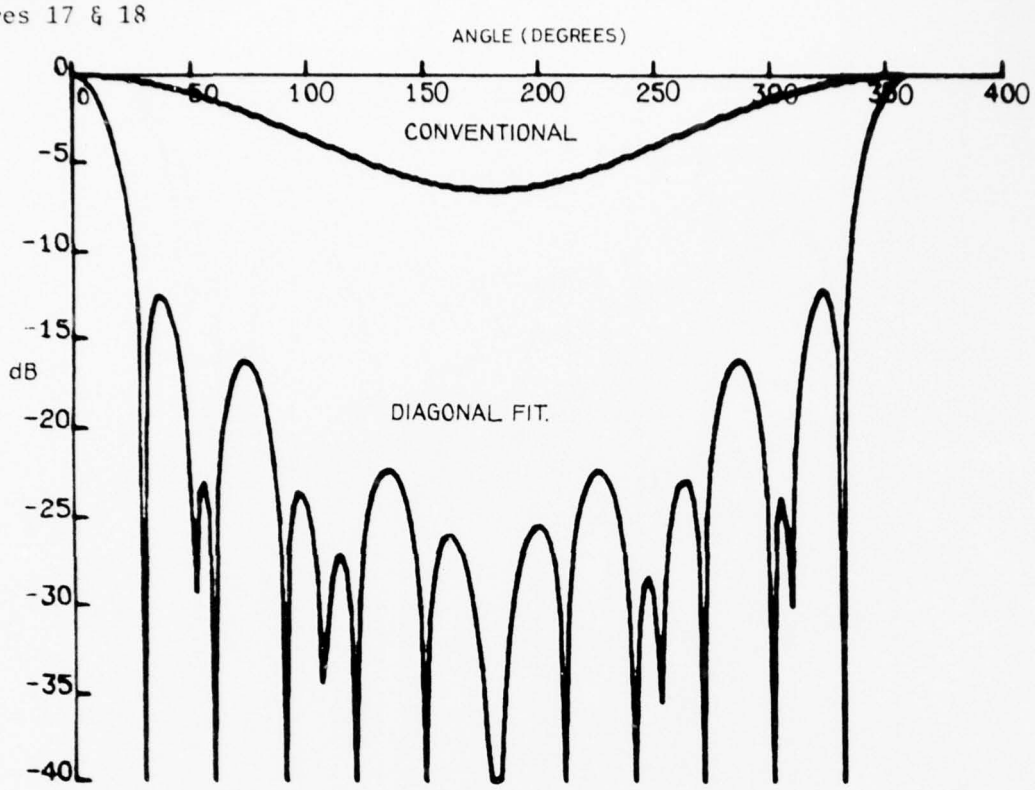


Figure 17. 10-element ring array - diagonal fit beamformer, $R/\lambda = 0.125$

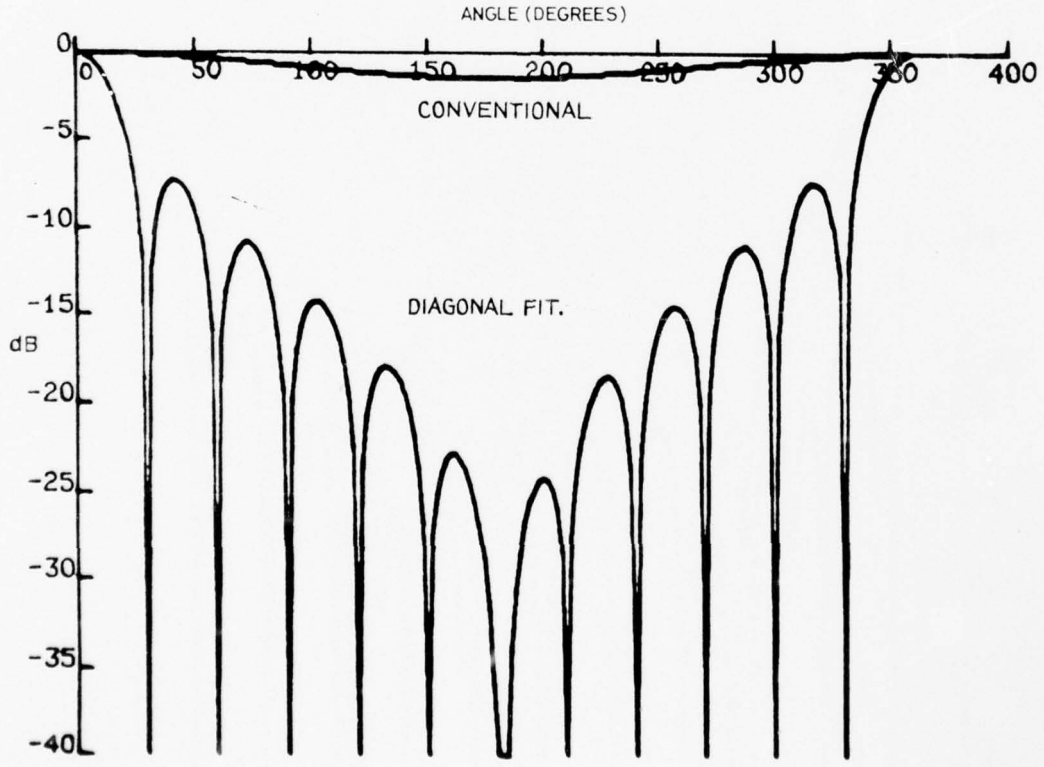


Figure 18. 10-element ring array - diagonal fit beamformer, $R/\lambda = 0.0625$

SUM AND SQUARE LAW SIGNAL PROCESSORS WITH MULTIPLE CLIPPED INPUTS

Prof. R.G. Keats, Department of Mathematics,
University of Newcastle, N.S.W.

S U M M A R Y

The recent work of Cheng and Stokes on the processing of clipped signals from two or three receivers is extended and generalised to n receivers and a number of the restrictions removed. In particular, the restriction that the noise in each receiver has the same normalised correlation function as the signal is removed.

The output signal to noise ratio for the clipped processor is compared with the unclipped case and it is shown that for large n this signal to noise ratio for the clipped processor exceeds that for the unclipped.

Mathematically - Plackett's result is used to expand the orthant probabilities involved in increasing powers of the input signal to noise ratio.

1. INTRODUCTION

The application of statistical hypothesis testing to the detection of signals in additive background noise is well established; see, for example, Helstrom [5], Van Trees [15], and more recent work by Kailath [6],[7],[8]. In many of these applications the outputs of a number of receivers are combined and processed to give a quantity on which a decision concerning the presence or absence of a signal may be based. This paper is concerned with sum and square law processors as illustrated in figure (1), and in particular, a comparison is made of the performance of two such processors; viz., the analogue sum and square law processor (A.S.S.P.), and the polarity sum and square law processor (P.S.S.P.).

Early work in this field was described by Faran and Hills [4] and other authors; for example, [3],[9],[12],[14]. This work was extended by Cheng[2], and both he and Stokes gave extensive resumés of this and related work in their theses, Cheng [1] and Stokes [13].

The measure, used by most authors, of the performance of processors similar to those discussed in this paper, is the ratio of signal power to noise power at the output of the processor. The signal power is defined as the square of the difference between the mean output in the presence of signal and the mean output in the absence of signal. The noise power is defined as the variance of the output in the presence of signal. Other measures of performance are, of course, possible, but as mentioned above, this signal to noise ratio is widely used, and accordingly it will also be used here.

The operation of the class of processors discussed here is shown in figure (1). The inputs to each receiver consist of a signal S plus noise N_i; these inputs are assumed to be realisations of stationary random processes. Although the signals at each receiver are identical realisations of the same process, it is possible that they are shifted in time relative to each other. It will be assumed that this shift is not present or, alternatively, delays are inserted to synchronise the signals from each receiver. Following the time delay the inputs to each receiver may be transformed linearly or non-linearly and the members of the class of processors studied here differ only in the type of transformation used at this point. In the case of the A.S.S.P. the transformation is the identity, whereas the P.S.S.P. incorporates the transformation

$$A_i = \mathrm{sgn}[I_i],$$

where

$$I_i = S+N_i$$

and the function sgn is defined by

$$\mathrm{sgn}(x) = 1, \quad x \geqslant 0,$$

$$\mathrm{sgn}(x) = -1, \quad x < 0.$$

Although one of the two cases compared in this paper uses the identity transformation, any linear transformation may be studied similarly and with very little added complication, since the relation between the statistics of a normal input to a linear filter and its output is well known; see, for example, Laning and Battin [10].

Following these transformations, the outputs A_i corresponding to each receiver are summed and squared to give the quantity $X = [\Sigma A_i]^2$ which is then integrated using the weighting function

$$W(t) = \frac{1}{T} \exp\left(\frac{-t}{T}\right)$$

to give the processor output

$$Y(t) = \frac{1}{T} \int_0^\infty e^{-\frac{\mu}{T}} X(t-\mu)\,d\mu .$$

The following additional assumptions will be made.

(a) The signal S(t) and the noise N(t) present at each receiver are stationary, normal, random processes with zero mean, variances σ_S^2, σ_N^2, and the same normalised correlation functions $\rho(\tau)$.

(b) All cross-correlation functions are identically zero.

(c) The input signal to noise ratio is small. This further assumption is not restrictive in most, if not all, practical cases.

2. PREVIOUS RELEVANT RESULTS

The following results will be used in this paper without derivation. They have been quoted and used by many authors and can all be found in Stokes [13].

$$E[Y] = E[X] = E\left(\sum_{i=1}^{n} A_i\right)^2 = \sum_{i=1}^{n} E[A_i^2] + \sum_{i\neq j}^{n} E[A_i A_j], \qquad (1)$$

$$\sigma_Y^2 = \frac{1}{T}\int_0^\infty e^{-\frac{\mu}{T}} \rho_X(\mu)\,d\mu, \qquad (2)$$

where

$\rho_X(\mu)$ is the covariance function of X.

If we define B_i to be the quantity A_i delayed by μ, then

$$\begin{aligned}\rho_X(\mu) &= E[(\Sigma A_i)^2(\Sigma B_j)^2] - E^2(\Sigma A_i)^2 \\ &= E[(\Sigma A_i^2)(\Sigma B_j^2)] + 2E[(\Sigma A_i^2)(\Sigma' B_i B_j)] \\ &\quad + 2E[(\Sigma B_i^2)(\Sigma' A_i A_j)] + 4E[\Sigma' A_i A_j B_i B_j] \\ &\quad + 4E[\Sigma' A_i A_j B_j B_k] + 4E[\Sigma' A_i A_j B_k B_\ell] - E^2[\Sigma A_i^2 + 2\Sigma' A_i A_j] \end{aligned} \qquad (3)$$

where Σ' indicates that the summation is to be taken over all terms such that $i < j$, $k < \ell$ and all distinct indices take different values.

The nub of the problem thus becomes the calculation of this correlation function $\rho_X(\mu)$.

In the case of the A.S.S.P. the following results are fairly trivial extensions of those appearing in the literature [13]. For this case the following results apply.

(a) In the absence of signal

$$E[Y(t)] = E[X(t)] = n\sigma_n^2$$

(b) In the presence of signal

$$E[Y(t)] = E[X(t)] = n^2\sigma_S^2+n\sigma_N^2,$$

while the autocorrelation function

$$\rho_X(\tau) = 2\rho^2(\tau)[n^2\sigma_S^2+n\sigma_N^2]^2.$$

The signal to noise ratio for this processor may then be written

$$\frac{n^2 a^2}{2K[1+(n-1)a]^2},$$

where

$$a = \frac{\sigma_S^2}{\sigma_S^2+\sigma_N^2},$$

and

$$K = \frac{1}{T}\int_0^\infty \exp(-\frac{\mu}{T})\rho^2(\mu)\,d\mu.$$

Cheng [1] has solved the problem of the P.S.S.P. in the case of two inputs. His result can be written

$$SNR_{PSSP(2)} = \frac{(\arcsin a)^2}{\frac{\pi^2}{16T}\int_0^\infty \exp(-\frac{\mu}{T})\rho_X(\mu)\,d\mu},$$

$$\rho_X(\mu) = \frac{16}{\pi^2}\{[\arcsin \rho(\mu)]^2-[\arcsin a\rho(\mu)]^2\}.$$

Stokes endeavoured to extend Cheng's results to n = 3 but was unable to find a closed form expression for the term $E[\Sigma' A_i A_j B_j B_k]$ in $\rho_X(\mu)$, (3). He obtained numerical values by a most complicated computational programme. This difficulty again arose when he endeavoured to extend the problem to higher values of n, but as can be seen from (3) no further difficulties arise. In fact, Stokes' expression for $\rho_X(\mu)$ [13, p.152] becomes

$$2n(n-1)E[A_iA_jB_iB_j]+4n(n-1)(n-2)E[A_iA_jB_jB_k]$$

$$+n(n-1)(n-2)(n-3)E[A_iA_jB_kB_\ell]-\frac{4}{\pi^2}n^2(n-1)^2[\arcsin(a)]^2. \qquad (4)$$

Closed form expressions are known [12] in this case for $E[A_iA_jB_iB_j]$ and $E[A_iA_jB_kB_\ell]$, but after a long search no closed form could be found for $E[A_iA_jB_jB_k]$.

These three expectations all depend on the evaluation of orthant quadrivariate probabilities; i.e., the evaluation of fourfold integrals of the form

$$\left(\frac{1}{2\pi}\right)^2 |Q|^{\frac{1}{2}} \int_0^\infty\int_0^\infty\int_0^\infty\int_0^\infty \exp -\tfrac{1}{2}(\underset{\sim}{x}^T Q \underset{\sim}{x})\, d\underset{\sim}{x}, \qquad (5)$$

where $Q = R^{-1}$ and R is the correlation matrix of four jointly normal variates having zero mean and unit variance. For the three cases above, the matrix R takes the form

$$\begin{pmatrix} 1 & a & b & -ab \\ a & 1 & ab & -b \\ b & ab & 1 & -a \\ -ab & -b & -a & 1 \end{pmatrix}, \qquad (6)$$

$$\begin{pmatrix} 1 & a & ab & ab \\ a & 1 & b & ab \\ ab & b & 1 & a \\ ab & ab & a & 1 \end{pmatrix} \qquad (7)$$

and

$$\begin{pmatrix} 1 & a & ab & -ab \\ a & 1 & ab & -ab \\ ab & ab & 1 & -a \\ -ab & -ab & -a & 1 \end{pmatrix} \qquad (8)$$

The one involving matrix (7) proved intractable.

3. OUTPUT SIGNAL TO NOISE RATIO OF THE P.S.S.P. (n)

Progress in this P.S.S.P. problem became possible when attempts to obtain a closed form expression for all three orthant probabilities were abandoned and a new approach adopted. Realising that the cases of practical interest were confined to those where the signal to noise ratio, and therefore the parameter a, were small, it was decided to look for a series expansion of expressions like (5).

To avoid confusion in notation, we will take advantage of the fact that correlation matrices are symmetric and write the general orthant probability for four variates as

$$F(\rho_{12},\rho_{13},\rho_{14},\rho_{23},\rho_{24},\rho_{34}),$$

i.e., F is a function of the six correlation coefficients of the four variates involved. In particular cases we will also use notations such as $\phi(a,b)$ so that, for example, in the case corresponding to the matrix (7)

$$F(a,ab,ab,b,ab,a) = \phi(a,b).$$

Expanding $\phi(a,b)$ at a = 0,

$$\phi(a,b) = \phi(0,b) + a\left(\frac{\partial\phi}{\partial a}\right)_{a=0} + \frac{a^2}{2!}\left(\frac{\partial^2\phi}{\partial a^2}\right)_{a=0} + + \tag{9}$$

and the problem reduces to evaluating the expressions ϕ, $\frac{\partial\phi}{\partial a}$, $\frac{\partial^2\phi}{\partial a^2}$, etc., at a = 0. Plackett's result [11] may be applied to the evaluation of these expressions, since in general

$$\left(\frac{\partial\phi}{\partial a}\right)_{a=0} = \left(\Sigma\frac{\partial F}{\partial\rho_{ij}}\frac{\partial\rho_{ij}}{\partial a}\right)_{a=0},$$

with more complicated, but still manageable expressions, for higher derivatives. Expressions for $\frac{\partial\rho_{ij}}{\partial a}$ are trivial while expressions such as $\frac{\partial^2 F}{\partial\rho_{ij}\partial\rho_{k\ell}}$ follow from [11].

For example, following Plackett [11], we have

$$\frac{\partial^2 F}{\partial\rho_{ij}\partial\rho_{k\ell}} = \left(\frac{1}{2\pi}\right)^2 |Q|^{\frac{1}{2}} \int_0^\infty\int_0^\infty\int_0^\infty\int_0^\infty \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}\frac{\partial}{\partial x_k}\frac{\partial}{\partial x_\ell} \exp -\tfrac{1}{2}(\underset{\sim}{x}^T Q \underset{\sim}{x})\, d\underset{\sim}{x} \tag{10}$$

and the value of this expression at a = 0 is easily obtained in the following way. Since a is not involved in either the integrations or the differentiations we are only concerned with the matrix $Q_o = Q$ evaluated at a = 0; e.g., in the example (7),

$$Q_o^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & b & 0 \\ 0 & b & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Furthermore, if all the indices i,j,k,ℓ are distinct it is easily seen that

$$\frac{\partial^2 \phi}{\partial \rho_{ij} \partial \rho_{k\ell}} = \left(\frac{1}{2\pi}\right)^2 |Q_o|^{\frac{1}{2}} ,$$

$$= \frac{1}{4\pi^2} \sqrt{1-b^2}$$

Similarly, if i = k = 2, j = ℓ = 3, we have

$$\left(\frac{\partial^2 \phi}{\partial \rho_{23}^2}\right)_{a=0} = \left(\frac{1}{2\pi}\right)^2 |Q_o|^{\frac{1}{2}} \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \left(\frac{\partial}{\partial x_2}\right)^2 \left(\frac{\partial}{\partial x_3}\right)^2 \exp -\tfrac{1}{2}(\underset{\sim}{x}^T Q_o \underset{\sim}{x}) d\underset{\sim}{x},$$

$$Q_o = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{1-b^2} & \frac{-b}{1-b^2} & 0 \\ 0 & \frac{-b}{1-b^2} & \frac{1}{1-b^2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

So that

$$\left(\frac{\partial^2 \phi}{\partial \rho_{23}^2}\right)_{a=0} = \left(\frac{1}{2\pi}\right)^2 |Q_o|^{\frac{1}{2}} \int_0^\infty \int_0^\infty \frac{b}{1-b^2} \exp[-\tfrac{1}{2}(x_3^2 + x_4^2)] \, dx_3 \, dx_4$$

$$= \frac{b}{8\pi (1-b^2)^{3/2}} .$$

4. DISCUSSION AND AN EXAMPLE

The partial derivatives required in order to evaluate $\frac{\partial \phi}{\partial a}$, $\frac{\partial^2 \phi}{\partial a^2}$ and $\frac{\partial^3 \phi}{\partial a^3}$ at a = 0 for the three cases corresponding to the matrices (6), (7), (8), have been calculated and appear in the appendix. The expectations required for the evaluation of the expression (4) are also derived in the appendix and evaluated using these orthant probabilities, but neglecting terms of order a^4 or higher. The signal to noise ratio of the A.S.S.P.$_{(n)}$ is

$$\frac{n^2 a^2}{2K[1+(n-1)a]^2} \qquad (11)$$

as given in Section 2. The signal to noise ratio of the P.S.S.P.$_{(n)}$, after neglecting terms of order a^4 and higher, may be written in the form

$$\frac{n(n-1)a^2}{\alpha n^2+\beta n+\gamma}, \tag{12}$$

where:

$$\alpha = 2a^2K(1-2a),$$

$$\beta = 4aI_1+a^2(4I_3-10K-4)+a^3(\tfrac{2}{3}I_2-4I_4+20K),$$

$$\gamma = 2I_5-8aI_1+a^2(10K-8I_3+8)+a^3(8I_4-\tfrac{4}{3}I_2-24K)$$

The quantity K was defined in Section 2.

The integrals I_1 to I_5 are of the form

$$I_1 = \frac{1}{T}\int_0^\infty \exp(-\tfrac{\mu}{T})\, b \arcsin(b)\, d\mu,$$

$$I_2 = \frac{1}{T}\int_0^\infty \exp(-\tfrac{\mu}{T})\, b^3 \arcsin(b)\, d\mu,$$

$$I_3 = \frac{1}{T}\int_0^\infty \exp(-\tfrac{\mu}{T})\,(1-b^2)^{\frac{1}{2}}\, d\mu,$$

$$I_4 = \frac{1}{T}\int_0^\infty \exp(-\tfrac{\mu}{T})\,(1-b^2)^{\frac{1}{2}} b^2\, d\mu,$$

$$I_5 = \frac{1}{T}\int_0^\infty \exp(-\tfrac{\mu}{T})\,[\arcsin(b)]^2\, d\mu.$$

All these integrals may be evaluated once the correlation function $b = \rho(\mu)$ is decided.

The expression (12) has been evaluated for a range of values of n using the parameter values,

$$a = \frac{1}{16}, \qquad b = \rho(\mu) = e^{-2|\mu|}$$

and for T = both 5 and 100. The results appear in decibel form in figures (2) and (3). The quotient of the signal to noise ratio of the A.S.S.P.$_{(n)}$ divided by that of the P.S.S.P.$_{(n)}$ appears as figure (4); the graph for T = 5 has been reproduced here; that for T = 100 is very little different.

As would be expected, the performance of each processor improves as the number of receivers, n, increases. More importantly, however, the performance of the P.S.S.P.$_{(n)}$ relative to the A.S.S.P.$_{(n)}$ improves as n increases. The most surprising result, (cf. e.g., [12]) however, is that for large n the performance of the P.S.S.P.$_{(n)}$ is, using this signal to noise criterion, better than that of

the A.S.S.P.$_{(n)}$. In fact, as $n \to \infty$ the quotient mentioned above approaches 1-2a. A similar result noted by Cheng [1] has been dismissed since, unlike the present case, it appeared to apply only for large values of the input signal to noise ratio. An examination of the expressions (11) and (12) shows that, in this present case also, the phenomenon could have been expressed in terms of a; 1-2a becomes smaller as a increases. However, the present work is only valid for small a, although it is probable that increasing a does have a similar effect to increasing n.

It has now been found that the approach described in Section 3 can be applied to the general case; i.e., when the random processes describing the signal and noise are allowed to have differing variances and correlation functions. Work on this general case has been completed and will shortly be submitted for publication.

ACKNOWLEDGEMENT

This work was carried out under a grant to the University of Newcastle from the Department of Defence, Australia.

REFERENCES

No.	Author	Title
[1]	Cheng, M.C.	"On a class of non-linear transformations of Gaussian random processes". Ph.D. Thesis, University of Adelaide (1968).
[2]	Cheng, M.C.	"The clipping loss in correlation detectors for arbitrary input signal to noise ratios". I.E.E.E. Trans. on Information Theory, IT-14, 382-389 (1968).
[3]	Ekre, H.	"Polarity coincidence correlation detection of a weak noise source" I.E.E.E. Trans. on Information Theory, IT-9, 18-23 (1963).
[4]	Faran, J.J. Jr. and Hills, R. Jr.	"The Application of Correlation Techniques to Acoustic Receiving Systems". (Technical Memo No.28, Acoust. Res. Lab., Harvard, 1952).
[5]	Helstrom, C.W.	"Statistical Theory of Signal Detection". (London, Pergamon Press, 2nd ed. 1960).
[6]	Kailath, T.	"A General Likelihood Formula for Random Signals in Gaussian Noise". I.E.E.E. Trans. on Information Theory, IT-15, 350-361 (1969).
[7]	Kailath, T.	"The Innovations Approach to Detection and Estimation Theory". Proc. I.E.E.E., Vol. 58, 680-695 (1970).
[8]	Kailath, T.	"Some Extensions of the Innovations Theorem". B.S.T.J. Vol. 50, 1487-1494 (1971).
[9]	Kanefsky, M.	"Detection of weak signals with polarity coincidence arrays". I.E.E.E. Trans. on Information Theory, IT-12, 260-268 (1966).
[10]	Laning, J.H. and Battin, R.H.	"Random Processes in Automatic Control". (McGraw-Hill, 1956).
[11]	Plackett, R.L.	"A reduction formula for normal multivariate integrals". Biometrika, 41, 351-360 (1954).
[12]	Schultheiss, P.M. and Tuteur, F.B.	"Optimum and suboptimum detection of directional Gaussian signals in an isotropic Gaussian noise field, Part II: degradation of detectability due to clipping". I.E.E.E. Trans. on Military Electronics, MIL-9, 208-211 (1965).

No.	Author	Title
[13]	Stokes, B.J.	"Signal detection using infinite clipping: A survey and an application". B.Math. Thesis, University of Newcastle (1973).
[14]	Thomas, J.B. and Williams, T.R.	"On the detection of signals in non-stationary noise by product arrays". Journal of Acoust. Soc. of America, Vol. 31, 453-462 (1959).
[15]	Van Trees, H.L.	"Detection, Estimation and Modulation Theory - Part III: Radar-Sonar Signal Processing and Gaussian Signals in Noise". (New York, Wiley, 1971).

APPENDIX

DERIVATION OF THE REQUIRED EXPECTATIONS

In this appendix the three expectations required for the calculation of the correlation function $\rho_\chi(\mu)$ are determined after neglecting terms of a^4 or higher. Previous work by Cheng [1] is used to simplify the calculation of $E[A_iA_jB_iB_j]$.

In all three cases the required expectation may be written in a form similar to

$$E[A_iA_jB_iB_j] = 2P[A_iA_jB_iB_j = 1] - 1,$$

but for this case we note, [1],

$$E[A_iA_jB_iB_j] = \frac{4}{\pi^2}\{[\arcsin(a)]^2 + [\arcsin(b)]^2 - [\arcsin(ab)]^2\}$$

$$= \frac{4}{\pi^2}\{[\arcsin(b)]^2 + a^2(1-b^2) + 0(a^4)\}.$$

$$E[A_iA_jB_jB_k] = 4[\phi_1(a,b) + \phi_1(a,-b) + \phi_2(a,b) + \phi_2(a,-b)] - 1$$

where

$$\phi_1(a,b) = F[a,ab,ab,b,ab,a],$$

$$\phi_2(a,b) = F[-a,ab,-ab,-b,ab,-a].$$

Using the results appearing in Table 1, this expectation can be written

$$\frac{2}{3\pi^2}\{6ab\arcsin(b) + 6a^2(1-b^2)^{1/2} + a^3[b^3\arcsin b - 6b^2(1-b^2)^{1/2}] + 0(a^4)\}.$$

Similarly

$$E[A_iA_jB_kB_\ell] = 1 - 16F[a,ab,-ab,ab,-ab,-a]$$

$$= \frac{4a^2}{\pi^2}(1+2b^2) - \frac{16a^3b^2}{\pi^2} + 0(a^4).$$

TABLE 1. PARTIAL DERIVATIVES OF ORTHANT PROBABILITIES NEEDED FOR THE CALCULATION OF THE REQUIRED EXPECTATIONS

$$\phi_1(a,b) = F(a,ab,ab,b,ab,a)$$

$$\phi_2(a,b) = F(-a,ab,-ab,-b,ab,-a)$$

The derivatives of $\phi_1(a,b)$ evaluated at $a = 0$ are given below, those for $\phi_2(a,b)$ may be obtained by changing the sign of b. Only those derivatives unequal to zero are given.

$$\phi_1(0,b) = \tfrac{1}{4}\left[\tfrac{1}{4} + \frac{1}{2\pi} \arcsin b\right]$$

$$\left(\frac{\partial \phi_1}{\partial \rho_{ij}}\right)_{a=0}$$

i j	Value
12,13,24,34	$\frac{1}{8\pi}$
14	$\frac{1}{2\pi}\left[\tfrac{1}{4}+\frac{1}{2\pi} \arcsin b\right]$

$$\left(\frac{\partial^2 \phi_1}{\partial \rho_{ij} \partial \rho_{k\ell}}\right)_{a=0}$$

i j k ℓ	Value
1234, 1324	$\left(\frac{1}{2\pi}\right)^2 (1-b^2)^{-\frac{1}{2}}$
1224, 1334	$\left(\frac{1}{2\pi}\right)^2 (-b)(1-b^2)^{-\frac{1}{2}}$

$$\left(\frac{\partial^3 \phi_1}{\partial \rho_{ij} \partial \rho_k \partial \rho_{mn}}\right)_{a=0}$$

ij kℓ mn	Value
121212, 131313, 242424, 343434	$\frac{1}{8\pi}$
121214, 131314, 142424, 143434	$\frac{b}{4\pi^2}(1-b^2)^{-\frac{1}{2}}$
121314, 142434, 212324, 132334	$\frac{-1}{4\pi^2}(1-b^2)^{-\frac{1}{2}}$
141414, 232323	$\frac{1}{8\pi^2}(\pi+2 \arcsin b)$

Defining $\phi_3(a,b) = F[a,ab,-ab,ab,-ab,-a]$ as required for $E[A_i,A_j,B_k,B_\ell]$ the required partial derivatives are found by equating b = 0 in those for $\phi_1(a,b)$.

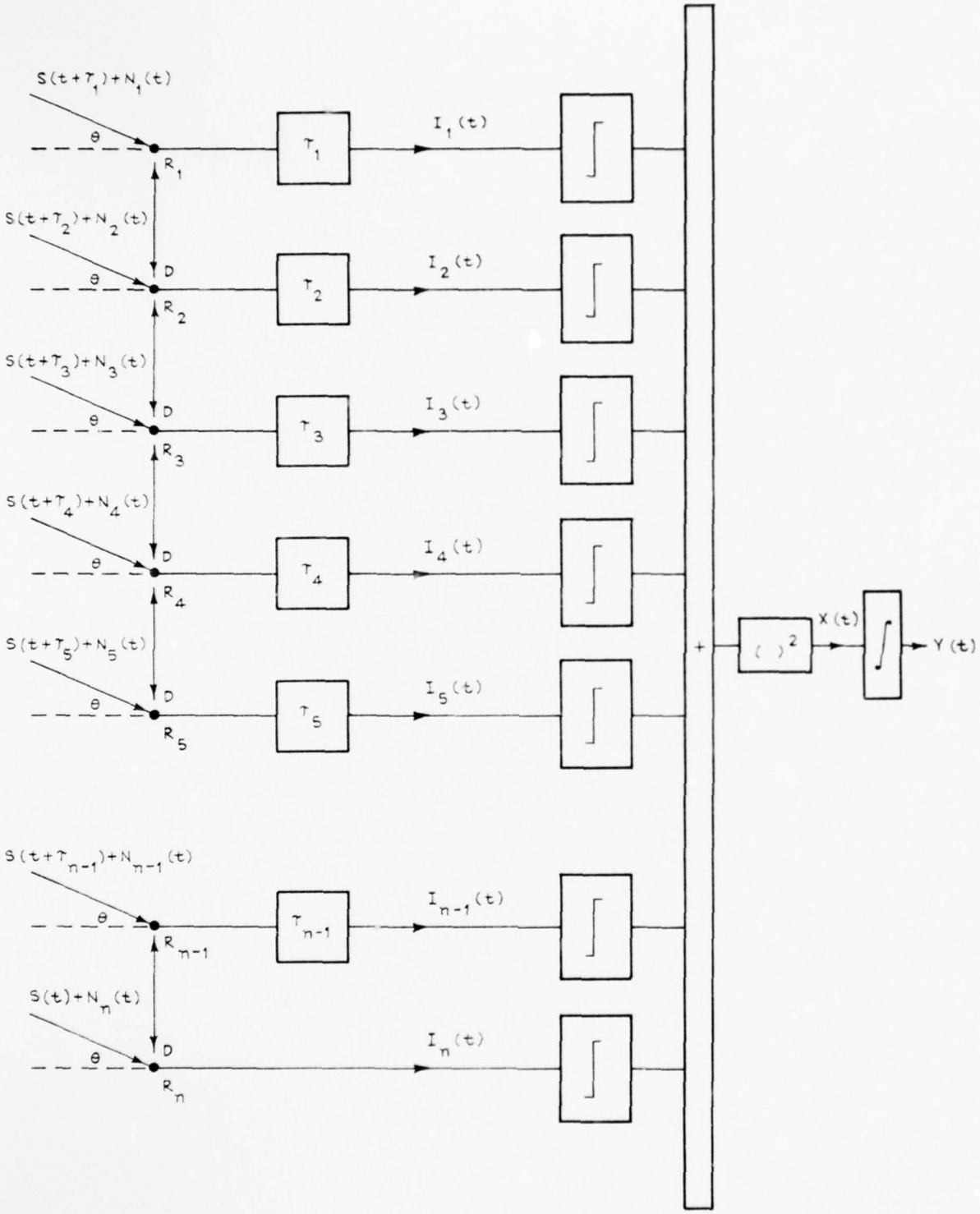


Figure 1. A sum and square law processor

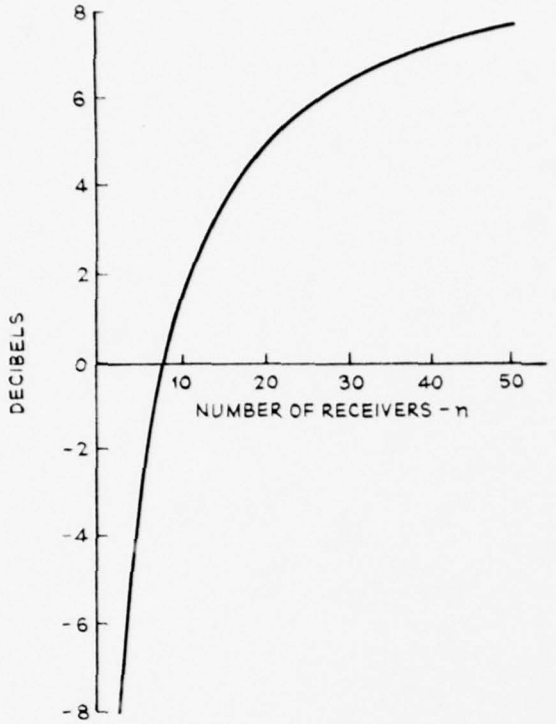


Figure 2. Output signal to noise ratio of clipped sum and square law processor for T = 5

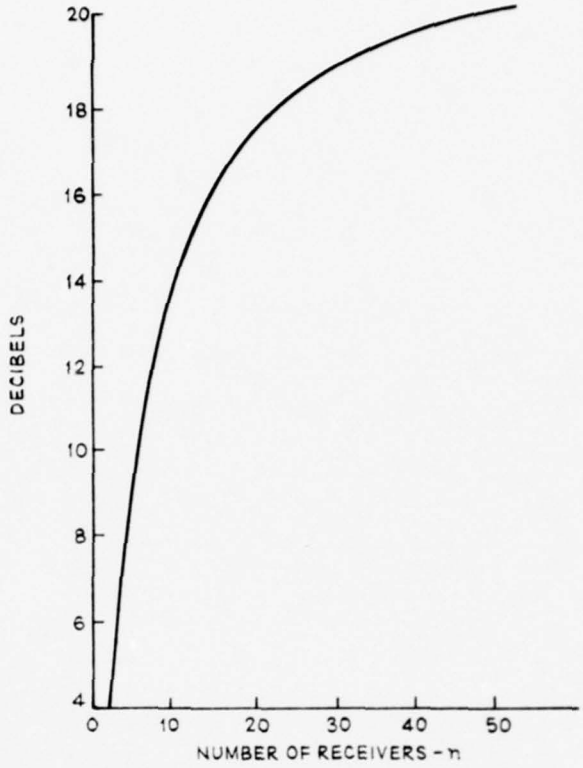


Figure 3. Output signal to noise ratio of clipped sum and square law processor for T = 100

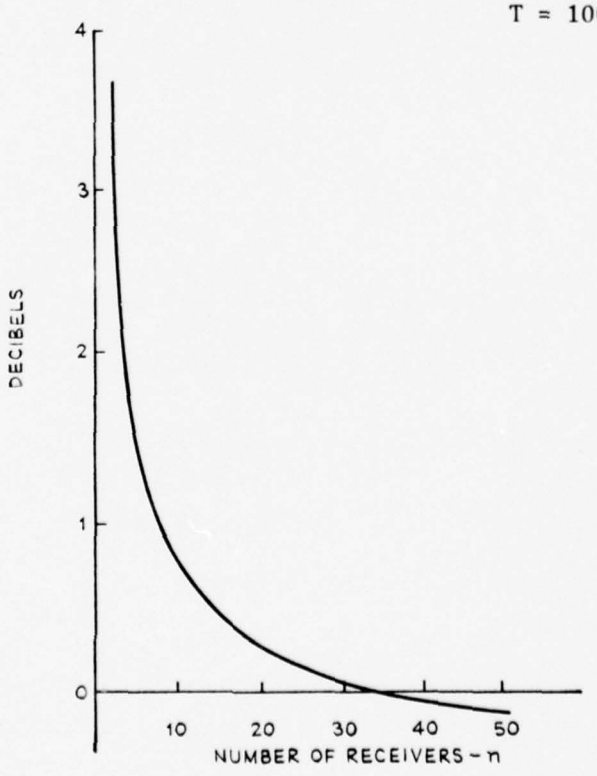


Figure 4. Comparison of output signal to noise ratio of analogue and clipped sum and square law processors for T = 5

MEASURING THE DIRECTIONS OF INTERFERING SIGNALS USING SMALL ARRAYS

J.D. Henstridge, Department of Statistics,
Research School of Social Sciences, Australian National University

S U M M A R Y

Most techniques for processing signals from small arrays which give consistent estimates of signal direction fail to give useful results when more than one signal is present. In this paper a pseudo-maximum likelihood method is used for the problem of two independent stochastic signals in the presence of noise which need not be Gaussian. Consistent estimators are given for the cases where the signal spectra are known and where they are not known. Simulation results are given to show small sample properties of the estimators.

1. INTRODUCTION

The purpose of this paper is to discuss estimators of signal directions and other parameters such as velocity, dispersion, etc., when the signal is propagated as a wave and observed by an array of separate receivers. The emphasis is on small arrays which would not be able to resolve signals from different directions when used in the conventional sense. The small size will also allow reasonably advanced processing of the outputs. The statistical treatment of large arrays which are usually of regular design is very different (Hinich and Shaman, 1972).

The one signal problem with small arrays has been treated comprehensibly by Hannan (1975). However the two signal case is considerably more difficult, especially if the signals cannot be separated (e.g., if they are not in disjoint frequency bands). Deconvolution techniques (e.g., Fujinawa, 1974 or McDonough, 1975) give reasonable results for extended sources but tend to behave badly with point sources and the results are difficult to access statistically.

We shall be concerned with estimators using the discrete time series obtained by sampling the output of each receiver in the array at regular intervals. Without loss of generality, we shall take this interval to be unity, giving us time series with spectra on $[-\pi,\pi]$. The discrete sampling can give a problem of aliasing whereby high frequencies are confused with lower ones (see Koopmans, 1974). This must be overcome by filtering out the high frequencies before the sampling process as the estimators we propose can be badly affected by them.

The estimators will be given for a sample of N observations in time. In practice N will be chosen to be highly composite and preferably a power of 2 so that the Fast Fourier Transform (FFT) can be used. Without the FFT, the methods proposed would be impractical for any but small sample sizes.

Further, we shall restrict the analysis to the case where the signals are independent of each other and of the noise, and the noise at each receiver has the same spectrum as and is independent of the noise at all other receivers. The receivers are assumed to be omnidirectional. Each of these conditions could be relaxed but only with an increase in complexity and without leading to significantly different results.

Only two signals will be considered in this paper but nothing in the analysis will be unique to the two signal problem.

The model for the output of the a-th receiver at time t is

$$y(t,a) = z_1(t,a) + z_2(t,a) + x(t,a), \quad a = 1,\ldots,p, \tag{1}$$

where the signals z_1 and z_2 are of the form

$$z_1(t,a) = \int \exp\{i[t\omega - \theta_1(a,\omega,\tau)]\}\, d\zeta_1(\omega) \tag{2(a)}$$

$$z_2(t,a) = \int \exp\{i[t\omega - \theta_2(a,\omega,\tau)]\}\, d\zeta_2(\omega) \tag{2(b)}$$

where ζ_1 and ζ_2 are random measures with zero mean and orthogonal increments as is usual and $x(t,a)$ is the noise. $\theta_1(a,\omega,\tau)$ and $\theta_2(a,\omega,\tau)$ are the phase delays of the signals at the receiver a and frequency ω and are a function of τ, an

unknown parameter vector, the true value of which (τ_0) must be estimated. In general, τ would be a vector in some subset of Euclidian space with components which specify, for instance, the direction or velocity of the two signals. Only the one parameter vector is used for both signals since they may have some properties in common, such as velocity. An example of such a θ would be

$$\theta(a,\omega,\tau) = \omega \underline{\xi}(a) \cdot \underline{u}/c(\omega)$$

where $\underline{\xi}(a)$ is the position of the receiver a, $\underline{u}$ is the unit direction vector of the signal and a function of τ and $c(\omega)$ is the velocity of transmission at frequency ω.

2. ONE SIGNAL CASE

For the one signal case, the almost universally used estimator can be expressed as that τ which maximises the function

$$q_N(\tau) = N^{-1} \sum_s W(\omega_s) \left| \sum_a Y_a(\omega_s)\, e^{i\,\theta(a,\omega_s,\tau)} \right|^2 \qquad (3)$$

where

$$Y_a(\omega) = N^{-\frac{1}{2}} \sum_n y(n,a)\, e^{in\omega} \qquad (4)$$

and $\omega_s = 2\,s/N$, $s = -N/2, \ldots, N/2$. The fourier transformation can be carried out using the FFT. The spectral weight function W is chosen so as to be in some sense optimal. It has been shown that this estimator is consistent (that is $\hat{\tau}_N \to \tau_0$ as $N \to \infty$ almost surely) under the very mild condition of ergodicity; under slightly stronger conditions, particularly upon the fourth-order moments of the signal, a central limit theorem holds, showing that $N^{\frac{1}{2}}(\hat{\tau}_N - \tau_0)$ tends to be normally distributed with zero mean and a specified variance (Hannan 1975). The signal and noise processes do not have to be Gaussian. The asymptotic variance depends only upon the second-order properties of the signal and noise. This is a point of some importance since a variance depending on higher-order moments is often useless because these moments are so difficult to measure in practice.

The derivation of equation (3) deserves some comment since it can provide a starting point for estimators of more than one signal. In one sense it is an obvious estimator obtained by the modelled phase differences between receivers upon the observed ones. It can also be found by a pseudo maximum likelihood technique. If the signal and noise are Gaussian, the $Y_a(\omega_s)$ are normally distributed and independent for different s, with zero mean and variance matrix $\hat{f}(\omega_s,\tau)$, the spectral matrix as predicted by the model with parameter τ. This gives a log likelihood of the form

$$\tfrac{1}{2}N_s^{-1} \det \hat{f}(\omega_s,\tau) - Y^*(\omega_s)\, \hat{f}^{-1}(\omega_s,\tau)\, Y(\omega_s) \qquad (5)$$

where Y is the vector with components Y_a. This procedure is outlined by Hannan (1970, p. 395).

For the one signal problem, the (a,b) entry of $\hat{f}(\omega,\tau)$ is $f_x(\omega)\,\delta_{ab} + f_z(\omega)\, e^{i[\theta(b,\omega,\tau) - \theta(a,\omega,\tau)]}$ where f_x is the spectrum of the noise, f_z is the spectrum of the signal and δ is the Kronecker delta function. $\det \hat{f}(\omega,\tau)$ can be shown to be independent of τ. After disregarding all other terms which are independent of τ, we arrive at equation (3) with $W(\omega) = f_x(\omega)^{-1}\{(f_x(\omega) + pf_z(\omega)\}^{-1}$. Although Gaussian assumptions are made in the derivation of the estimator, they are not necessary in its application. Experience indicates that this type of estimator is close to the best available in practice. What is more, it gives us an expression for the weight function W.

A third method of deriving equation (3) is to minimise the expression

$$\sum_s W(\omega_s) \sum_a \sum_b \left| I_{ab}(\omega_s) - \hat{f}_{ab}(\omega_s,\tau) \right|^2 \qquad (6)$$

where $I_{ab}(\omega) = \overline{Y_a(\omega)}\, Y_b(\omega)$ is the periodogram. After disregarding all terms not depending on τ, we arrive at equation (3). This method does not lead to the optimal weight function W.

3. TWO SIGNAL CASE

3.1 Stochastic signals

We now attempt to extend some of these ideas to a two signal situation. We shall consider only absolutely continuous spectra with densities f_1 and f_2 for the signals. To simplify, we shall set

$$\varphi(\omega,\tau) = \sum_a e^{i\{\theta_1(a,\omega,\tau) - \theta_2(a,\omega,\tau)\}}.$$

Where possible to do so without confusion, we shall drop references to ω and τ in the following.

The model spectral density f now has entries of the form

$$\hat{f}_{ab} = f_{x\,ab} + f_1 e^{i[\theta_1(b) - \theta_I(b)]} + f_2\, e^{i[\theta_2(b) - \theta_2(a)]}.$$

When this is substituted into equation (5) we obtain

$$q_N(\tau) = N^{-1} \sum_s \left\{ \log \Delta - f_x^{-1} \Delta^{-1} f_1 (f_x + pf_2) \left| \sum_a Y_a e^{i\theta_1(a)} \right|^2 \right.$$

$$+ f_2 (f_x + pf_1) \left| \sum_a Y_a e^{i\theta_2(a)} \right|^2$$

$$\left. - 2 f_1 f_2 R_e \left[\overline{\varphi} \sum_a Y_a e^{i\theta_1(a)} \sum_a Y_a e^{i\theta_2(a)} \right] \right\}$$

where

$$\Delta = f_x^2 + pf_x f_1 + pf_x f_2 + f_1 f_2 (p^2 - |\varphi|^2).$$

This can be shown to give reasonable results in practice but is hampered by several severe disadvantages.

(a) For consistency it requires knowledge of all spectra including their absolute levels.

(b) It is numerically rather complex leading to problems of numerical accuracy. In simulation studies, working at 9 significant digits it was found almost impossible to calculate the derivatives accurately enough to be usable. Working at a higher precision would so increase computation times as to make it of rather limited use.

(c) The asymptotic variance depends upon the fourth order cumulant.

If we substitute our expression for $\hat{f}$ into equation (6), we obtain

$$q_N(\tau) = N^{-1} \sum_s W(\omega_s) \left\{ f_1 \left| \sum_a Y_a e^{i\theta_1(a)} \right|^2 + \right.$$

$$\left. f_2 \left| \sum_a Y_a e^{i\theta_2(a)} \right|^2 - f_1 f_2 |\varphi|^2 \right\}. \qquad (8)$$

This estimator is much simpler and appears to overcome the numerical problems mentioned above. For very low signal levels, equation (8) approximates equation (7). It still has the other problems of equation (7).

Of these, the greatest handicap appears to be the requirement of knowing the spectra accurately. It would be impractical to consider the value of each spectral density a quantity to be estimated in each band, as this would lead to far too many parameters to be manageable. To model the spectra in some way so as to reduce the number of such quantities would often lead to inconsistent results if the model was chosen badly. The method of reducing the likelihood in equation (6) with respect to the spectra (that is, to replace the value of the spectral density for each band by that value, depending on τ which maximises the likelihood) is not practical due to the

restriction that the spectral densities remain non-negative.

3.2 Deterministic signals

A rather different approach is to consider the signals to be deterministic rather than stochastic, while keeping the noise stochastic. While this appears to be a big jump in terms of probability theory, it is hard to see the difference it makes in reality. If only one realization of a process is observed, it is impossible to decide whether the process is stochastic or deterministic. Possibly probability theory makes a distinction which may be misleading in practice.

The pseudo maximum likelihood approach lends itself to this treatment since, with only the noise being stochastic, the likelihood is expressed purely in terms of the distribution of the noise. The model for the signals is most easily presented in terms of their fourier transforms Z_{1a} and Z_{2a}. We let

$$Z_{1a}(\omega_s) = N^{-\frac{1}{2}} \sum_n z_1(n,a)\, e^{-i\omega_s n}$$

and

$$Z_{1a}(\omega_s) = Z_1(\omega_s)\, e^{-i\theta_1(a,\omega_s,\tau)}.$$

Z_{2a} is defined in exactly the same manner. The pseudo maximum likelihood estimator is then given by the maximum of

$$-\sum_s f_x^{-1} \sum_a \left| Y_a - Z_1 e^{-i\theta_1(a)} - Z_2 e^{-i\theta_2(a)} \right|^2, \tag{9}$$

the maximization to be taken over all τ, $Z_1(\omega_s)$, $Z_2(\omega_s)$. In a sense we are also estimating the phase and amplitude of each signal in each frequency band. However, we can reduce the likelihood with respect to Z_1 and Z_2 (since they are unrestricted in the complex plane) and we then obtain

$$q_N(\tau) = \sum_s f_x^{-1} \left\{ p \left| \sum_a Y_a e^{i\theta_1(a)} \right| 2 + p \left| \sum_a Y_a e^{i\theta_2(a)} \right|^2 \right.$$

$$\left. - 2\mathrm{Re}\left[\overline{\varphi} \sum_a Y_a e^{i\theta_1(a)} \sum_a Y_a e^{i\theta_2(a)} \right] \right\} (p^2 - |\varphi|^2)^{-1}. \tag{10}$$

This estimator appears to be the most useful one. Although it was derived using assumptions of deterministic signals and Gaussian noise, its application is far more general and it is easiest to examine its properties using stochastic signals.

(a) It does not depend upon knowledge of the signal spectra and for consistency does not require the noise spectrum either. However, it is more efficient if the noise spectrum is known.

(b) Numerically it is very stable.

(c) The asymtotic variance does not depend upon the higher-order moments of the signals or noise.

The proofs of consistency and the central limit theorem for this and the previous estimators will be available in a forthcoming paper.

There remain several interesting points about this estimator to be clarified. Even if one has estimates of the signal spectra, it would probably still be preferable to use equation (10) since it is consistent whether the estimates are good or not. It seems that the method for incorporating this extra information into equation (10) would be to replace the term f_x^{-1} by a different weight function.

4. DISCUSSION

There is a slight problem with all of these estimators in that the surface to be maximised has a large number of secondary maxima, analogous to the sidelobes of an array used in the conventional way. This is most noticeable when the signals have most of their power in a narrow spectral band and can be largely overcome by choosing the weight function so as to lessen the effects of these bands.

A few comments seem required on the extension of these estimators to more than two signals. It is obvious that estimators obtained by maximising equation (6) can handle an arbitrary number of signals for a given sized array. However there would be computational problems and with more signals the knowledge of the spectra would have to be still more accurate.

Estimators of the last type can only handle a maximum of p-1 signals, this limit being set by the number of degrees of freedom being used to estimate the amplitude and phase of each signal.

There seems to be an intermediate class of estimators or problems where the spectra are not known, but they can be assumed to be in some sense smooth. In this case the limit of p-1 could be passed. This would be of most use with two element arrays.

5. ACKNOWLEDGEMENT

I would like to acknowledge the assistance of Professor E.J. Hannan in the research leading to this paper.

REFERENCES

No.	Author	Title
1	Fujinawa, Y.	"Measurement of directional spectra of wind waves using an array of wave detectors". J. Oceanographical Soc. Japan 30, pp 10-22, 1974
2	Hannan, E.J.	"Multiple time series". Wiley, 1970 "Multivariate time series analysis". J. Multivariate Analysis, 3, pp 395-407, 1973 "Measuring the velocity of a signal". Contribution to "Perspectives in probability and statistics". (J. Gani, ed.). 1975
3	Hinich, M.J. and Shaman, P.	"Parameter estimation for an R dimensional plane wave observed with Gaussian errors". Annals Math. Stat. 43, pp 153-169, 1972
4	Koopmans, L.H.	"The spectral analysis of time series". Academic Press, 1974
5	McDonough, R.N.	"Spectral restoration for uniformly spaced arrays". J. Amer. Acoust. Soc. 57, pp 913-922, 1975

MAXIMUM ENTROPY SIGNAL PROCESSING

Dr. D.A. Gray, Weapons Research Establishment, Salisbury,
South Australia, 5108

S U M M A R Y

The theory of beamforming for a linear array with equispaced receivers can be shown to be equivalent to time series spectrum analysis. A new technique of spectrum analysis has been developed by J.P. Burg which estimates the data outside the observation interval whilst imposing the least constraints. This is known as the maximum entropy extension and may be applied to beamforming.

A theoretical derivation and discussion of the maximum entropy technique is given, with results of its application both to beamforming and time series analysis.

1. INTRODUCTION

In conventional spectrum analysis of a time series, the data are accumulated over a finite observation interval. An implicit assumption is usually made that values outside the observation interval are identically zero. This often limits the resolution of features of interest and distorts estimates of the power spectrum. A new technique for spectrum estimation has been proposed by J.P. Burg(ref.1,2) which overcomes these limitations by extrapolating either the data or the auto-correlation function outside the observation interval. The least constraints are imposed on this extrapolation and the resulting spectrum has been termed the maximum entropy (M.E.) spectrum.

A number of reports in the literature(ref.3,4,5) have demonstrated the superiority of the maximum entropy method (M.E.M.) over conventional techniques when the periodicities of interest are of the same order as the observation interval. These reports have used the M.E.M. for spectrum estimation. In addition, a technique(ref.2), based on minimizing the average power output of the associated prediction error filter, is used to estimate the autocorrelation function. In this paper the results are presented of an alternative application of the M.E.M. to acoustic data using a technique proposed by Radar(ref.6) to estimate the autocorrelation function.

For a linear array, frequency domain beamforming is essentially the formation of the wave-number spectrum, and for equi-spaced receivers the estimation of the wave-number spectrum is analogous to the estimation of the power spectrum of a time series. Some preliminary results from using this analogy to form the maximum entropy wave-number spectrum are discussed in the Section 4.

2. DERIVATION OF THE MAXIMUM ENTROPY SPECTRUM

The concept of entropy originated in thermodynamics as a measure of heat and then, more generally, as a measure of the disorder of a system. It was introduced into communication theory by Shannon(ref.7) who associated it with a measure of the information of a communication channel. Since entropy is a measure of disorder, the extrapolation chosen should maximize the entropy, thus ensuring that the least constraints consistent with the observed data have been imposed on the system. The application of this principle to spectrum analysis can be effected in two equivalent ways, since entropy can be defined in terms either of the autocorrelation function or of the spectrum. Throughout this section we assume that the sampled autocorrelation function has been estimated from the data, i.e., $r_j = \langle x_t \, x_{t-j} \rangle$ where x_i is the i-th sample and $\langle \ \rangle$ denotes some sort of averaging.

2.1 The maximum entropy extension of the autocorrelation function(ref.8)

Assuming the equisampled time series to be an N-th order Gaussian process, the entropy (H_N) is given by(ref.7)

$$H_N = \ln(2\pi e)^{\frac{N}{2}} |R_N|^{\frac{1}{2}},$$

where R_N, the Toeplitz covariance matrix of the time series, is given by

$$R_N = \begin{bmatrix} r_o & r_1 & \cdot & \cdot & \cdot & r_{N-2} & r_{N-1} \\ r_1 & r_o & r_1 & & & & r_{N-2} \\ \cdot & & \cdot & & & & \\ \cdot & & & \cdot & & & \\ \cdot & & & & \cdot & & \\ r_{N-2} & & & & & r_o & r_1 \\ r_{N-1} & r_{N-2} & & & & r_1 & r_o \end{bmatrix}$$

Let $\hat{r}_N$ be the extrapolation of the autocorrelation function by one lag. The expression for the entropy H_{N+1} is now a quadratic function of $\hat{r}_N$ and by choosing $\hat{r}_N$ so as to maximize H_{N+1}, i.e., setting $\frac{\partial H_{N+1}}{\partial \hat{r}_N} = 0$, a unique solution for the maximum entropy extension of the correlation sequence is obtained. Continuing this process for $\hat{r}_{N+1}$, $\hat{r}_{N+2}$, ..., etc., the autocorrelation function can be extended indefinitely and the M.E. spectrum obtained as the Fourier transform of this extended sequence.

The results of a comparison of the Fourier and maximum entropy extrapolations of 16 samples of the autocorrelation function $r_j = \frac{1}{\alpha} \delta_{j0} + \cos 2\pi \frac{j}{16}$, are shown in figure 1. This is the ideal correlation function of a 1 Hz sine wave in white noise with a signal-to-noise ratio of 10 log $\frac{\alpha}{2}$ dB. In all graphs, the spike at the origin due to the white noise component (i.e., $\frac{1}{\alpha} \delta_{j0}$) has been suppressed to improve the presentation. Figures 1(b), 1(c) and 1(d) are the M.E. extrapolations to 16 times the original time interval, with values for α of 10, 1 and 0.5 respectively. The maximum entropy extension appears to be modulated by an exponential decay whose time constant is inversely proportional to the signal-to-noise ratio.

2.2 Maximum entropy spectrum estimation

The maximum entropy spectrum may be obtained directly from the power spectrum S(f) by noting(ref.7) that the entropy H is given by the equation

$$H = \int_W \ln S(f)\, df$$

where W is the bandwidth.

It can be shown (see reference 9) that if H is maximized subject to the constraint that

$$\int_W S(f) e^{2\pi i f j \tau_o} df = r_j, \quad j = 0,1\ldots,N-1,$$

where τ_o is the sampling interval, then the M.E. spectrum is given by

$$S(f) = \frac{P_N}{\left| \sum_{j=0}^{N-1} \gamma_j \, e^{-2\pi i j \tau_o f} \right|^2} .$$

In the above equation $\gamma_o = 1$ and the remaining γ_j's and P_N are solutions of

$$\sum_{j=0}^{N-1} R_{ij} \, \gamma_j = P_N \, \delta_{i0},$$

$$\text{where } \underline{\delta}_{i0}{}^T = \{\delta_{i0}\} .$$

Since $\underline{R}_N$ is a Toeplitz matrix (i.e., $R_{ij} = r_{|i-j|}$), use may be made of efficient algorithms(ref.10) for the solution of this equation.

3. APPLICATION TO TIME SERIES ANALYSIS

To implement the techniques outlined above, an estimate of the autocorrelation lags $r_o, r_1, \ldots, r_{N-1}$ must be obtained from the data. Burg(ref.2) has proposed a technique which, by utilizing the relationship between maximum entropy spectrum analysis and the theory of predictive error filtering(ref.11), allows the γ's to be estimated directly. This technique has been successfully used and reported in the literature(ref.3). The technique uses the whole data sequence to estimate the γ_j's and the number of γ_j's estimated is recommended(ref.3) to be less than 15% of the data block. In applications of this type, however, relatively few data samples are in general used to estimate the power spectrum and consequently both the M.E.M. and conventional methods will have a large variance associated with the estimate of the power spectrum.

It is well known that if there is a large number of data samples, the variance of an estimator can usually be reduced by subdividing the data series into blocks and then averaging the estimates over these blocks. The remainder of this section will be devoted to demonstrating that the M.E.M. still shows considerable improvement over conventional analysis when used in this mode.

There are two main ways in which the M.E.M. can be applied to long, segmented data sequences. The first is to form the spectrum directly for each block by the methods indicated in reference 2 and then to obtain a resultant spectrum as the average of each block spectra. This approach is not considered in this paper. The approach adopted here is to estimate the autocorrelation function for each block and then to average these autocorrelations to obtain an improved statistical estimate of the autocorrelation function. The M.E. spectrum is then obtained from this autocorrelation function by the method described in Section 2.2.

Before describing the application of this to real data, some subtleties encountered in the estimation of the autocorrelation function will be described. The common method of zero filling each block, convolving it with itself and then averaging the resultant convolutions is shown in figure 2. However the autocorrelation function obtained by this method is not unbiased but is distorted by a triangular or Bartlett window function. The unsuitability of using this distorted autocorrelation function for the estimation of the M.E. spectrum is shown in figure 3. The M.E. spectrum resulting from applying a triangular window to the theoretical autocorrelation of a sine wave in white noise shows artifacts such as the splitting of spectral lines.

In order to overcome these difficulties, a technique proposed by Rader has been used to estimate the autocorrelation function. The method is indicated diagrammatically in figure 4 and involves the convolution of two contiguous blocks with the first of the blocks zero filled. This produces an unbiased estimator with the same variance for all lags. The convolution is efficiently implemented by use of the Fast Fourier Transform.

Comparisons of conventional and M.E. spectrum analyses using the techniques outlined above are shown in figures 5 to 9. In figure 5(a), 512 lags of the autocorrelation function of a 20 Hz sine wave mixed with band-limited white noise have been estimated from 5120 samples by splitting the data into 10 blocks. Using 64 lags of the autocorrelation function, the spectra in figures 5(b) and 5(c) were obtained. Relative to the height of the main peak the M.E.M. (figure 5(c)) has reduced the variance of the noise and eliminated the sidelobes associated with the Fourier estimate (figure 5(b)). The widths of the main peaks have also been decreased.

Figures 6 and 7 are the results of analysis of typical samples of noise radiated underwater by the propellers and machinery of ships. Using 20 blocks of 512 samples, 512 lags of the autocorrelation function have been estimates in figure 6(a) where a large spike at zero lag due to white noise has been reduced by a factor of five for presentation. Applying a Bartlett window to the full 512 lags in figure 6(a), the spectrum figure 6(b) was obtained. Using 128 lags of this correlation function the Fourier and maximum entropy spectra have been calculated and are shown in figures 7(a) and 7(b) respectively. Besides eliminating to a large extent the noise fluctuations and the ripple which are present in the Fourier spectrum, the M.E.M. has demonstrated quite clearly, by virtue of its greater resolution, that the 135 Hz spectral line is not a singlet but is in fact a doublet. Figure 7(c) shows that the conventional approach of applying a window (in this case a Bartlett one) to 128 lags in order to reduce the sidelobe structure so broadens the beamwidth that the splitting of the 135 Hz line is no longer discernible. As can be seen from figure 6(b) the splitting indicated by the M.E.M. is not an artifact.

In figure 8(a), 25 blocks each of 512 samples have been used to estimate 512 lags of a heavily modulated sine wave once again produced by shipping noise. Figure 8(b) is the conventional spectrum using the full 512 samples to indicate the modulation and additional lines. Using only 64 samples of the autocorrelation function, figures 9(a) and 9(b) demonstrate the ability of the M.E.M. to resolve the spectrum into the appropriate components. In particular, the M.E.M. has reduced the sidelobe structure so as to identify the lines between 200 and 600 Hz.

4. MAXIMUM ENTROPY BEAMFORMING AND ITS APPLICATION

For a linear array of N equi-spaced receivers (separated a distance d apart) an estimate of the frequency wave-number spectrum is obtained from(ref.12)

$$y(\theta, f) = \left\langle \left| \sum_{j=0}^{N-1} X_j(f) e^{-2\pi i k d} \right|^2 \right\rangle$$

where $k = \frac{f \sin \theta}{c}$, c is the propagation velocity and $X_j(f)$ is the output of the j-th receiver at a frequency f. Defining

$$r_j(f) = \sum_{i=0}^{N-j-1} \langle X_i(f) X^*_{j+i}(f) \rangle \qquad (1)$$

then

$$y(\theta, f) = \sum_{-N+1}^{N-1} r_j(f)\, e^{-2\pi i k d} .$$

The above is an implementation of a general result that, for a narrow-band plane wave-front signal, the wave-number spectrum is the Fourier transform of the autocorrelation function of the spatial distribution of the signal. In the above equation the assumption has been made that the autocorrelation is zero outside the array aperture. This is analogous to assuming in conventional Fourier analysis of time series that the autocorrelation function is zero outside the observation interval.

By exploiting the analogy between time series analysis and beamforming for a linear array, the principle of maximum entropy may be used to estimate the correlation sequences outside the array aperture. In terms of implementation of acoustic arrays, especially at low frequencies, if the M.E.M. can be shown to reliably extrapolate the data to, say, twice the array length, then this can offer considerable savings and benefits.

If the process whose wave-number spectrum is to be estimated can be assumed to be spatially stationary, i.e., the autocorrelation $r(x_i, x_j)$ is a function of $x_i - x_j$ only, then the M.E. wave-number spectrum for a line array of equi-spaced samples may be derived in a manner analogous to the time series case. If $r_p(f)$ is an estimate of the spatial correlation at a frequency f and a separation of pd, then the M.E. spectrum ($B(\theta, f)$) is given by

$$B(\theta, f) = \frac{P_N}{\left| \sum_{j=0}^{N-1} \gamma_j e^{-2\pi i k d} \right|^2} \qquad (2)$$

where $\underset{\sim}{R}\underset{\sim}{\gamma} = P_N \underset{\sim}{\delta}_{j0}$,

$$R_{ij} = r_{i-j} = R^*_{ji}, \qquad (3)$$

and

$$\underset{\sim}{\delta}^T_{j0} = \{\delta_{j0}\}.$$

An alternative formulation of the M.E. wavenumber spectrum is given in reference 13 where the assumption of spatial stationarity is not used. The spectrum in this case is also given by equations (2) and (3) but the matrix $\underset{\sim}{R}$ is no longer Toeplitz. It is important to realize the potential differences in an implementation of these two derivations. In the latter case R_{ij} is given by

$$R_{ij}(f) = \langle X_i^*(f) X_j(f) \rangle$$

whereas in the approach adopted in this section $R_{ij} = r_{i-j}$ is given by

$$r_j(f) = \frac{N}{N-j} \sum_{i=0}^{N-j-1} \langle X_i(f) X_{i+j}^*(f) \rangle . \qquad (4)$$

In the limiting case where $\langle\ \rangle$ represents ensemble averages and the X's are spatially stationary, then the two techniques are identical. In other cases the resulting M.E. spectra will in general be difficult.

Since the array is fixed in space the technique used in Section 3 cannot be used to estimate the spatial correlation function. Instead an estimate of the spatial correlation function is obtained using equation (1). The convolution expressed by equation (1) as discussed in Section 3 gives rise to a Bartlett window and is inherent in any conventional delay and sum beamformer.

One method of overcoming this difficulty is to remove the Bartlett weighting by premultiplying with the factor $\frac{N}{N-j}$ as shown in equation (4). Unfortunately this increases the variance of the estimates of $r_j(f)$ for j large in comparison with that for j relatively small and is probably the source* of the spikes which degrade the M.E. spectrum if the total number of lags of the correlation sequence is used. A brute force way of alleviating this problem is to use only a portion of the available lags in equation (2) to determine the prediction error coefficients.

As an example of these considerations the M.E. wave-number spectrum of a 32-element linear array has been calculated at a frequency corresponding to 0.56 λ. Twenty integrations have been effected to improve the estimate of $r_j(f)$ and the M.E. spectrum resulting from using 32, 16, 8 and 4 of the available 32 lags are shown in figure 10. The corresponding conventional spectrum with a Bartlett window using the full 32 lags is shown in figure 11. The improved resolution of the M.E.M. is apparent but unfortunately, as the number of available lags is increased, spurious peaks are introduced until, at 32 lags, the technique breaks down. Using only a portion of the available lags is not a wholly satisfactory approach and further work investigating the application of Burg's technique(ref.2) to wave-number spectrum and alternative approaches to the problem is presently being considered.

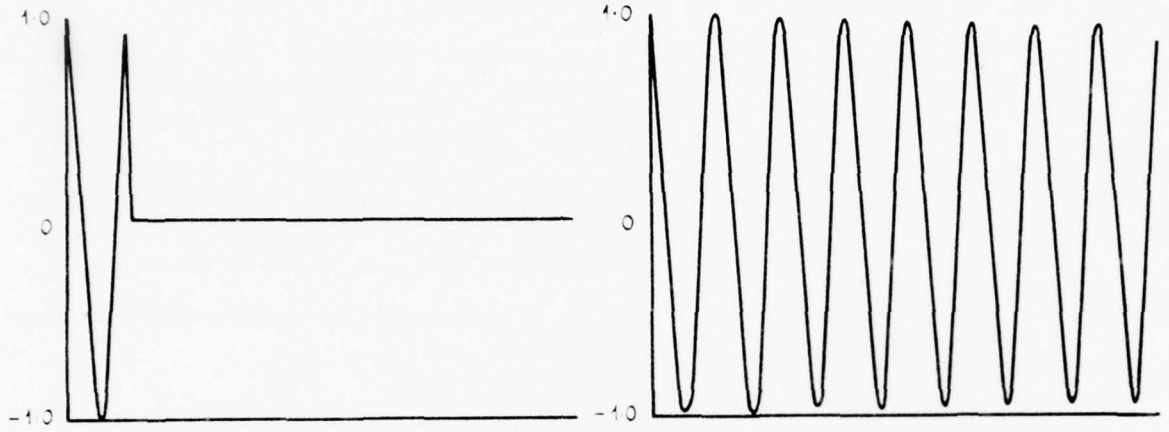
5. CONCLUDING REMARKS

The superiority of the M.E. technique over conventional analysis has been demonstrated for both spectrum analysis and beamforming. However, the advantages resulting from the M.E.M. may be eroded if particular care is not taken. Investigations into methods for understanding the difficulties associated with the M.E. technique are needed.

* The M.E.M. can be considered as a least-squares fit of an all-pole model to the data(ref.8) and as the number of lags of the autocorrelation function is increased then the poles increase in number and approach the unit circle(ref.5). The relatively large variance of some of the lags because of the proximity of the poles to the unit circle can then make the spectrum unstable.

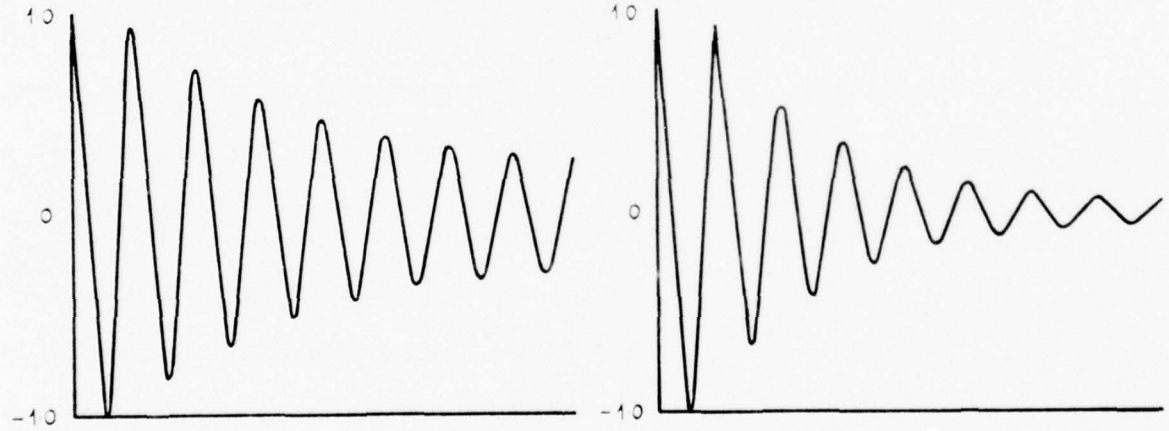
REFERENCES

No.	Author	Title
1	Burg, J.P.	"Maximum entropy spectrum analysis". Paper presented at 37th meeting of the Society of Exploration Geophysicists, Oklahoma City, Oklahoma (1967)
2	Burg, J.P.	"A new analysis technique for time series data". Paper presented at NATO Advanced Study Institute on Signal Processing, August 1968, Enschede, Netherlands
3	Ulrych, T.J.	"Maximum entropy power spectrum of truncated sinusoids". J. Geophys. Res. 77, 1396 (1972)
4	Jensen, O.G. and Ulrych, T.	"An analysis of the perturbations on Barnard's star". Astro. Jnl. 78, 1104 (1973)
5	Curie, R.G.	"Period and Q of the Chandler wobble". Geophys. J.R. astr. Soc. 38, 179 (1974)
6	Rader, C.M.	"An improved algorithm for high speed auto-correlation with applications to spectral estimation". IEEE Trans. Audio Electroacoust. AU-18, 439 (1970)
7	Shannon, C.E. and Weaver, W.	"The mathematical theory of communication". Urbana, I11: Univ. of Illinois Press, (1949)
8	A. van den Bos	"Alternative interpretation of maximum entropy spectral analysis". IEEE Trans. Info. Theory, 493 (1971)
9	Edwards, J.A. and Fitelson, M.M.	"Notes on maximum-entropy processing". IEEE Trans. Inform. Theory, 232 (1973)
10	Markel, J.D. and Gray, A.A.	"On autocorrelation equations as applied to speech analysis". IEEE Trans. Audio and Electroacoustics, AU-21, 69 (1973)
11	Ulrych, T.J., Smylie, D.E., Jensen, O.G. and	"Predictive filtering and smoothing of short records by using maximum entropy". Jnl. Geophys. Res., 78, 4959 (1973)
12	Rudnick, P.	"Digital beamforming in the frequency domain". J. Acoust. Soc. Amer. 44, 1454 (1968)
13	Barnard, T.E.	"Analytic studies of techniques for the computation of high-resolution wave-number spectra, Advanced array research, Special Report No.9". Texas (1969)



(a) Fourier

(b) $\alpha = 10$



(c) $\alpha = 1$

(d) $\alpha = 0.1$

Figure 1. Extrapolation of sine wave in white noise

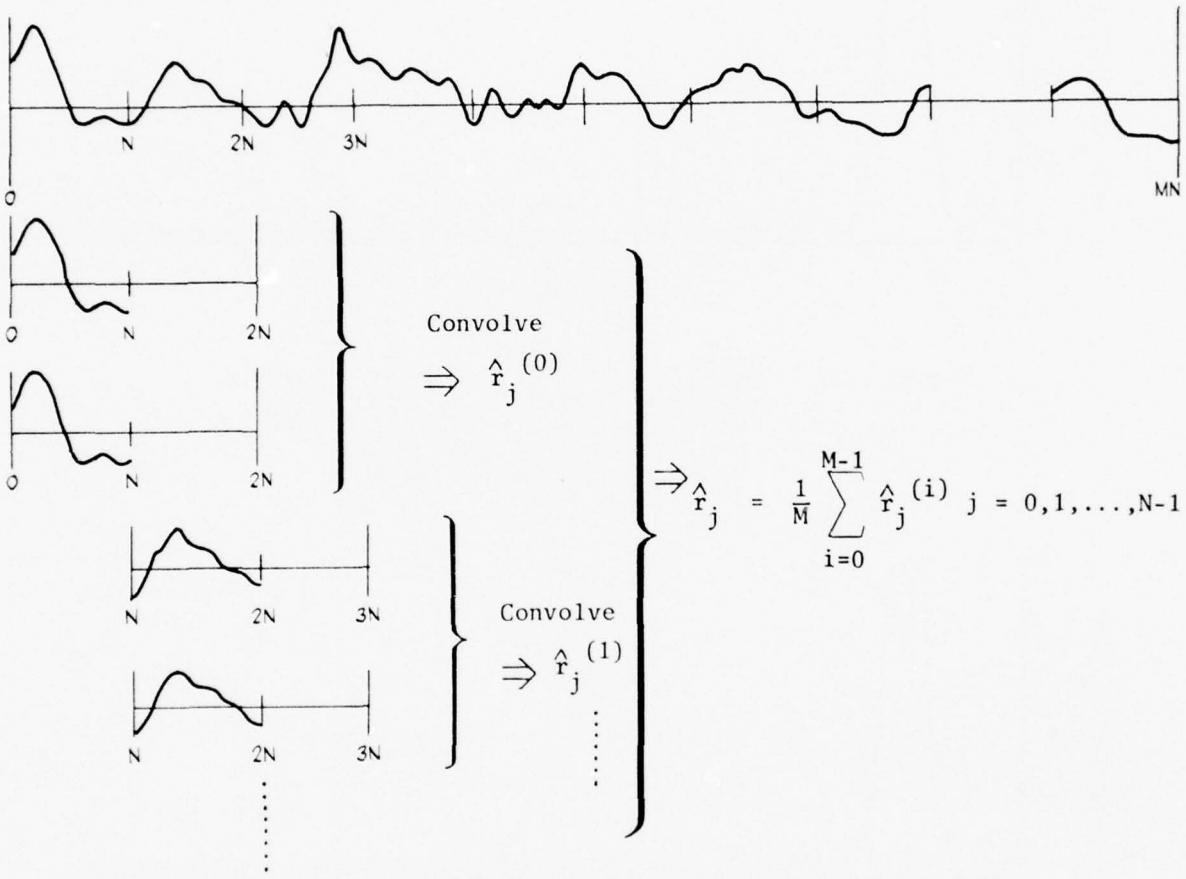


Figure 2. Segmenting the data and estimation of the correlation sequence with a triangular weighting

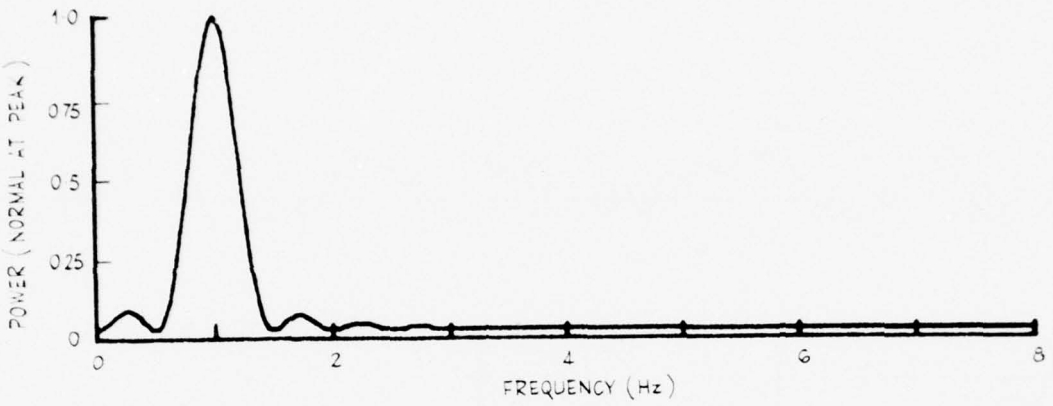


Figure 3(a). Bartlett spectrum of 1 Hz sine wave in white noise

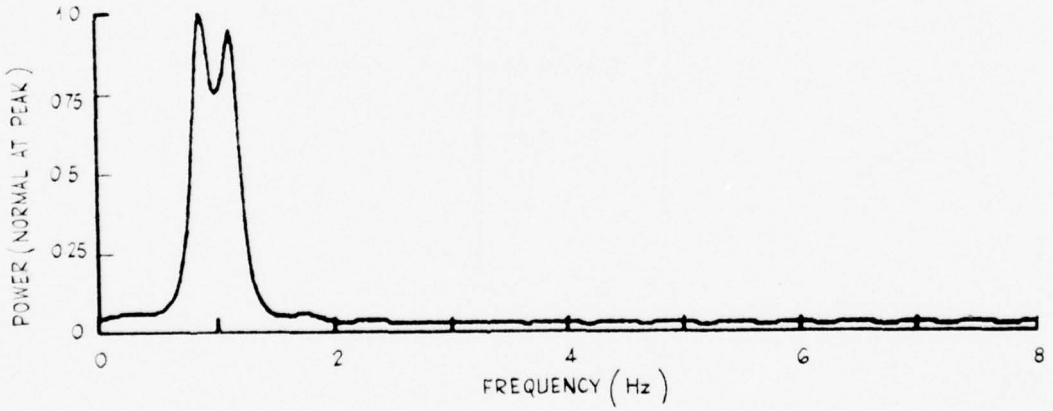


Figure 3(b). Maximum entropy of 1 Hz sine wave in white noise with a window

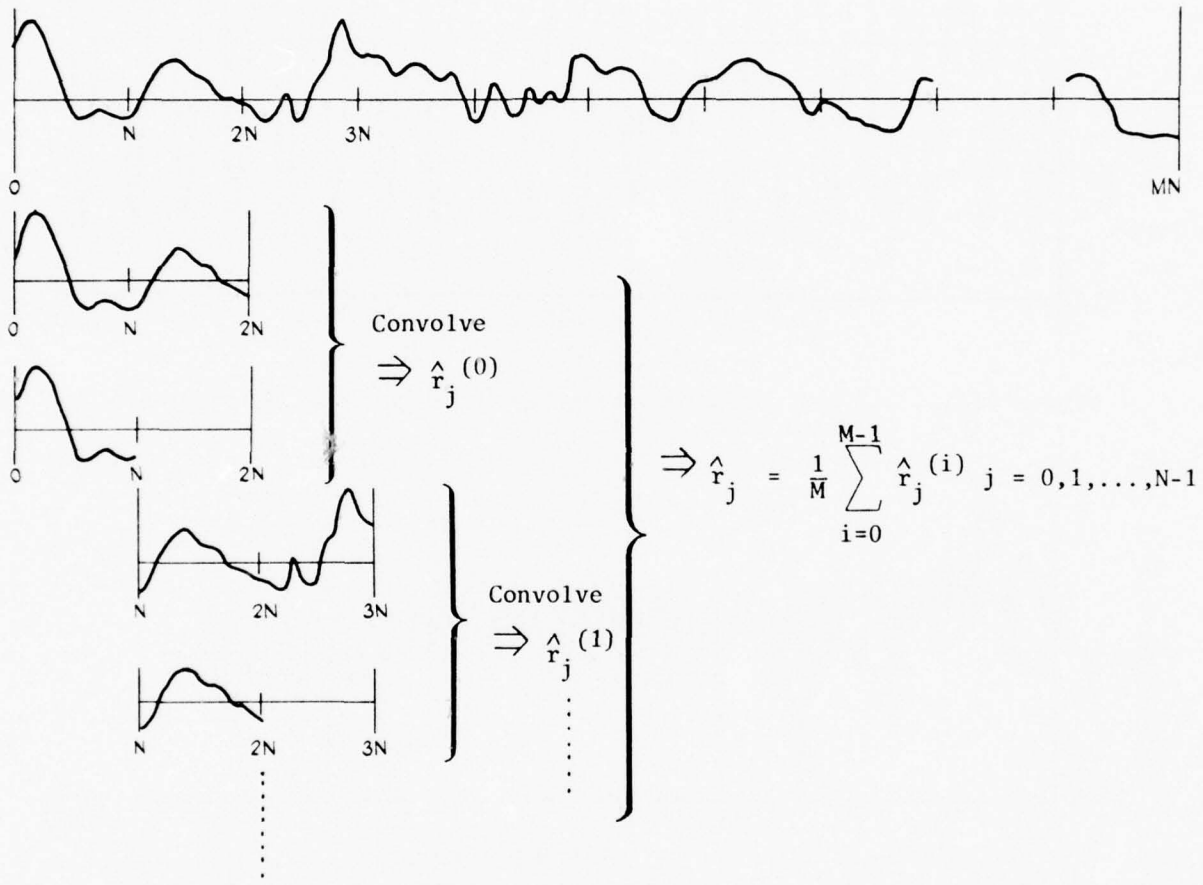


Figure 4. Raders technique for estimating the correlation sequence

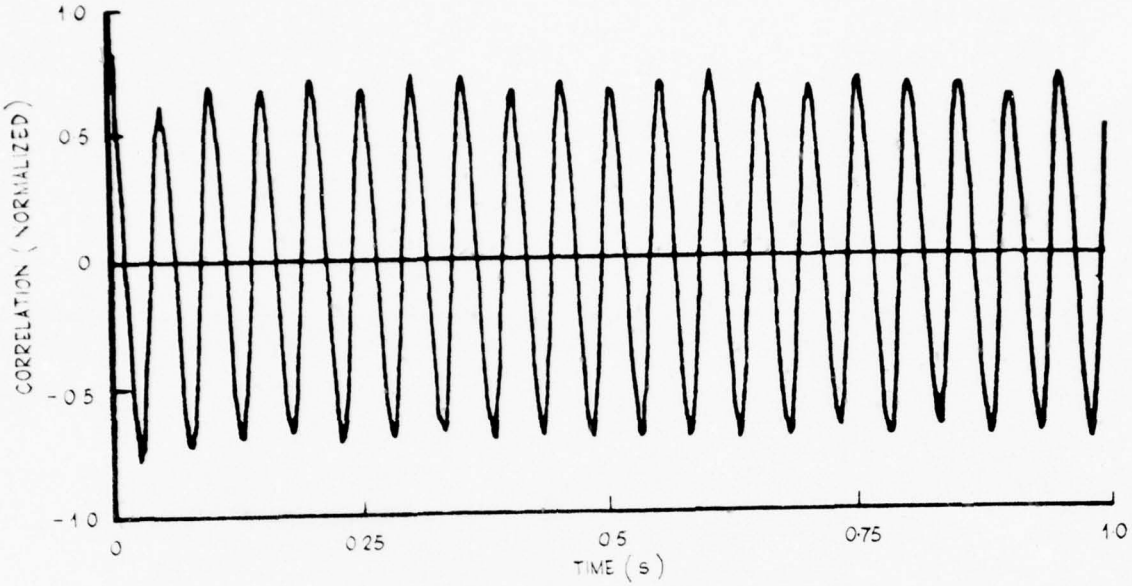


Figure 5(a). 512 lags of autocorrelation function of 20 Hz sine wave in white noise

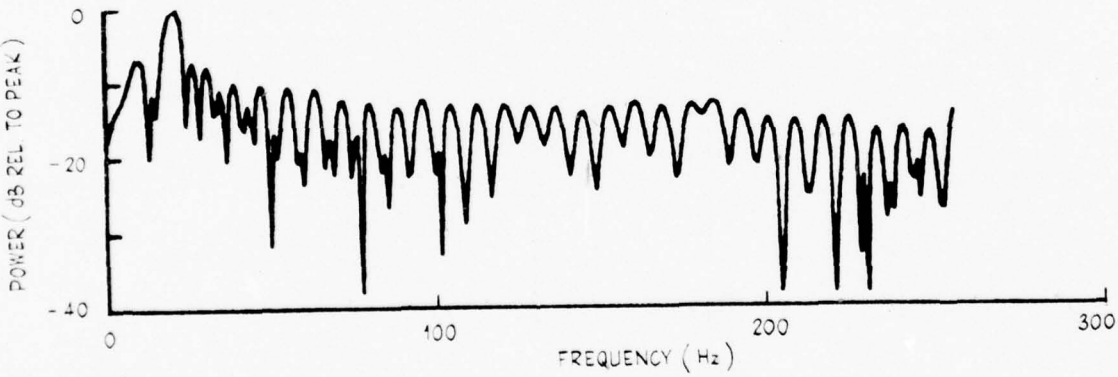


Figure 5(b). Fourier spectrum using 64 lags of figure 5(a)

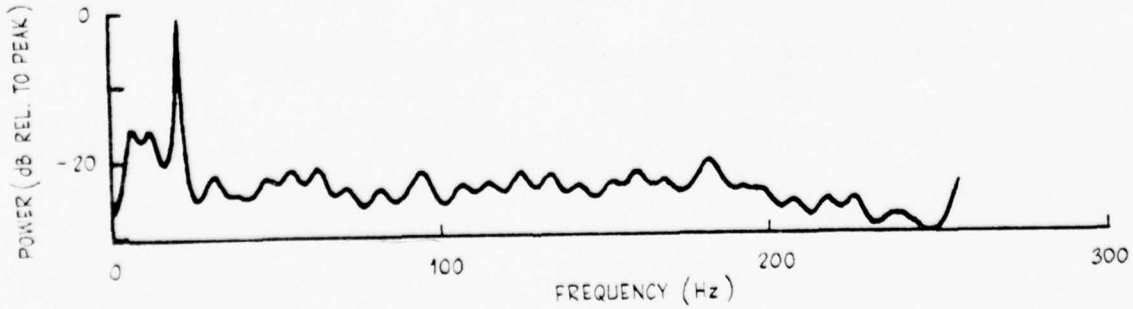


Figure 5(c). Maximum entropy spectrum using 64 lags of figure 5(a)

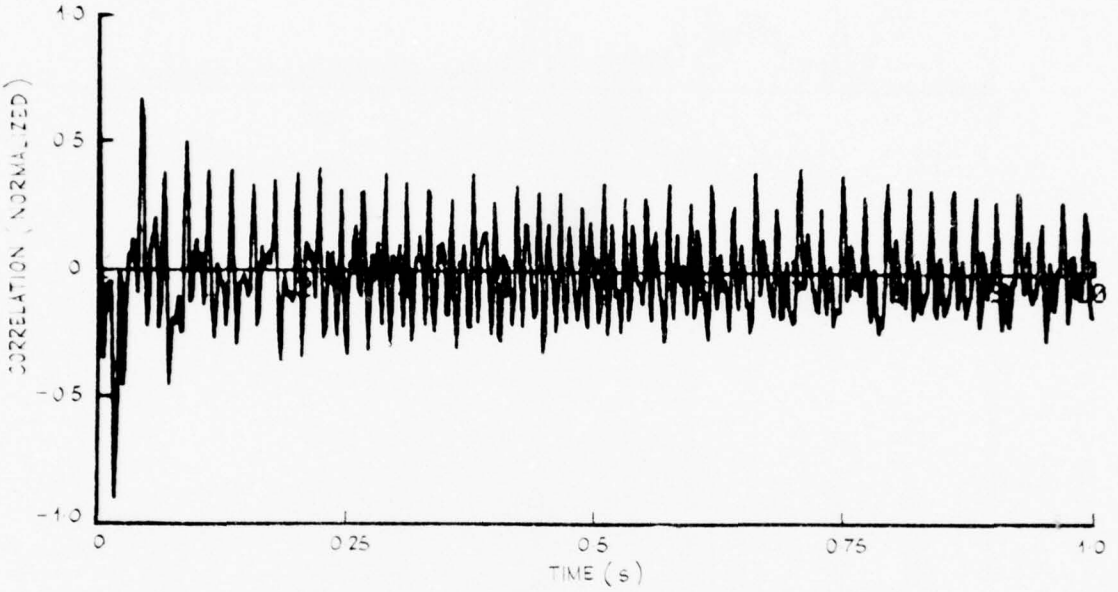


Figure 6(a). Autocorrelation (512 lags) of ship noise

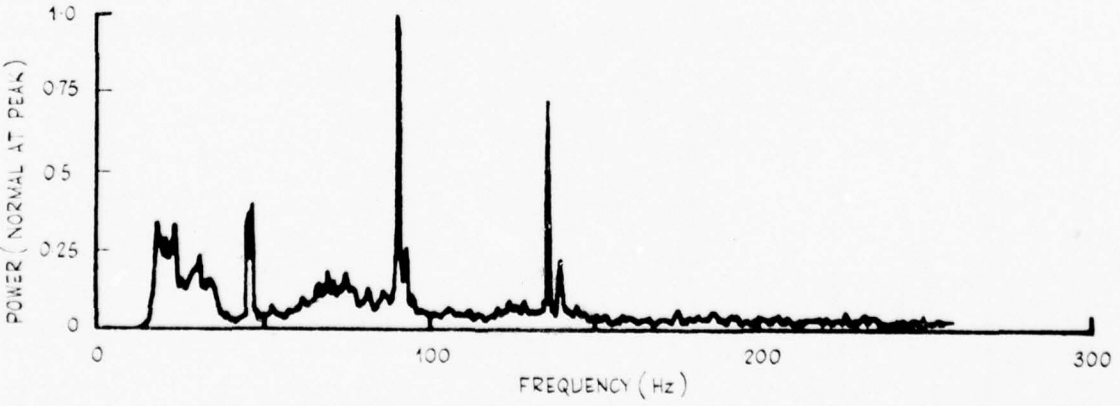


Figure 6(b). Bartlett spectrum corresponding to above

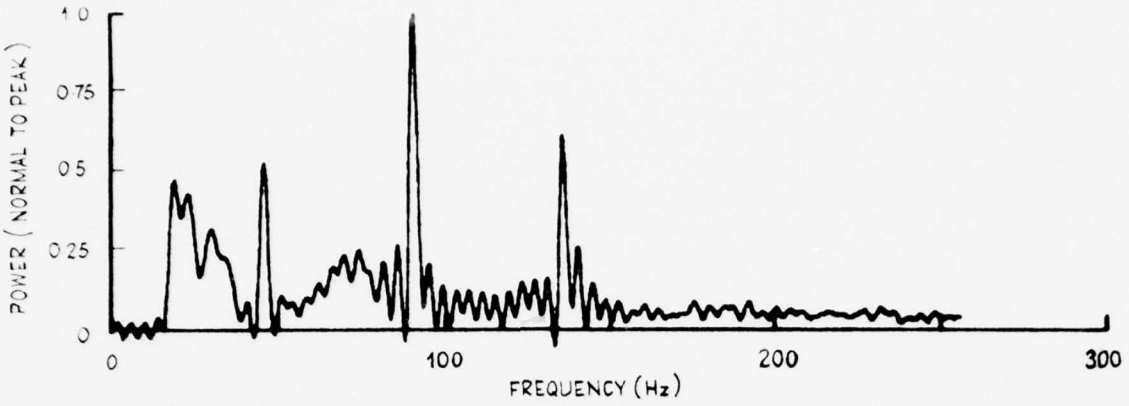


Figure 7(a). Fourier spectrum using 128 lags of figure 6(a)

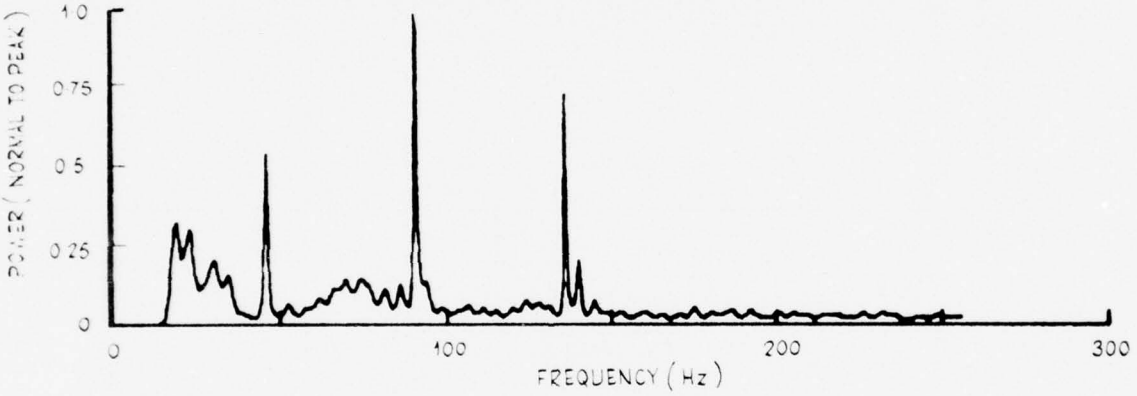


Figure 7(b). Maximum entropy spectrum using 128 lags of figure 6(a)

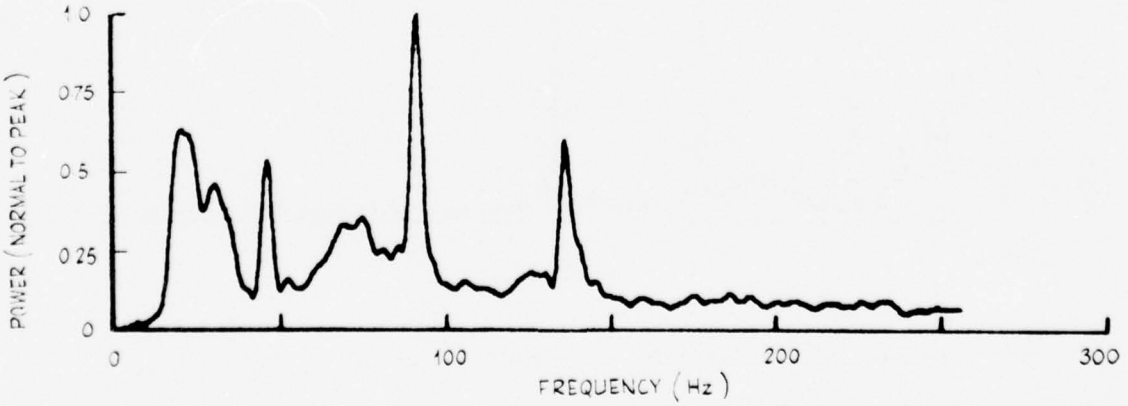


Figure 7(c). Bartlett spectrum using 128 lags of figure 6(a)

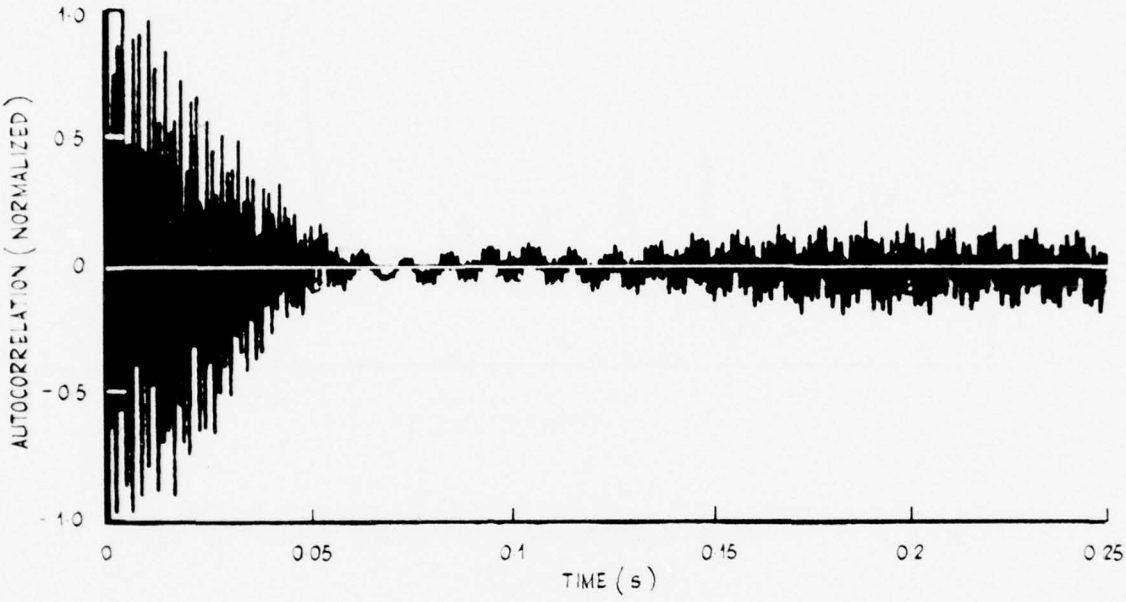


Figure 8(a). Autocorrelation function of ship noise (512 lags)

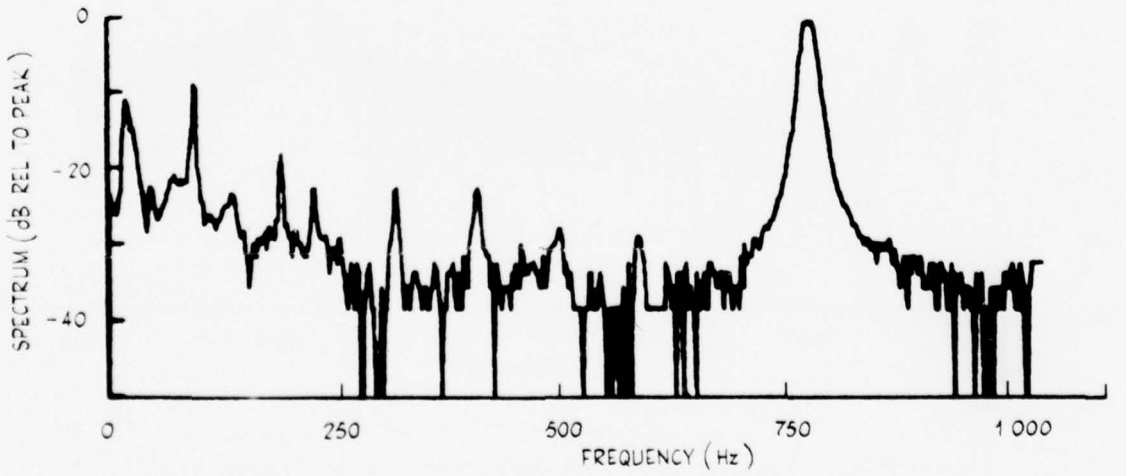


Figure 8(b). Bartlett spectrum using 512 lags of figure 8(a)

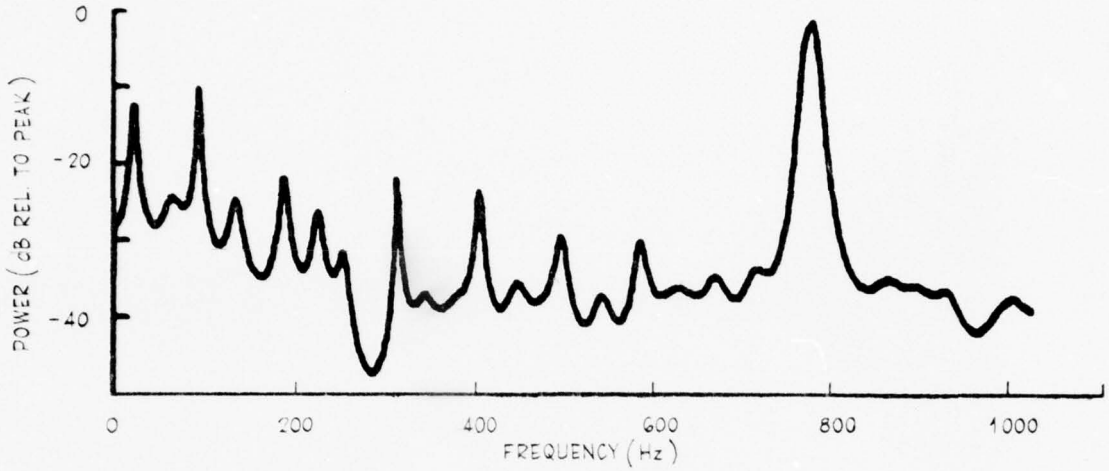


Figure 9(a). Maximum entropy spectrum using 64 lags of figure 8(a)

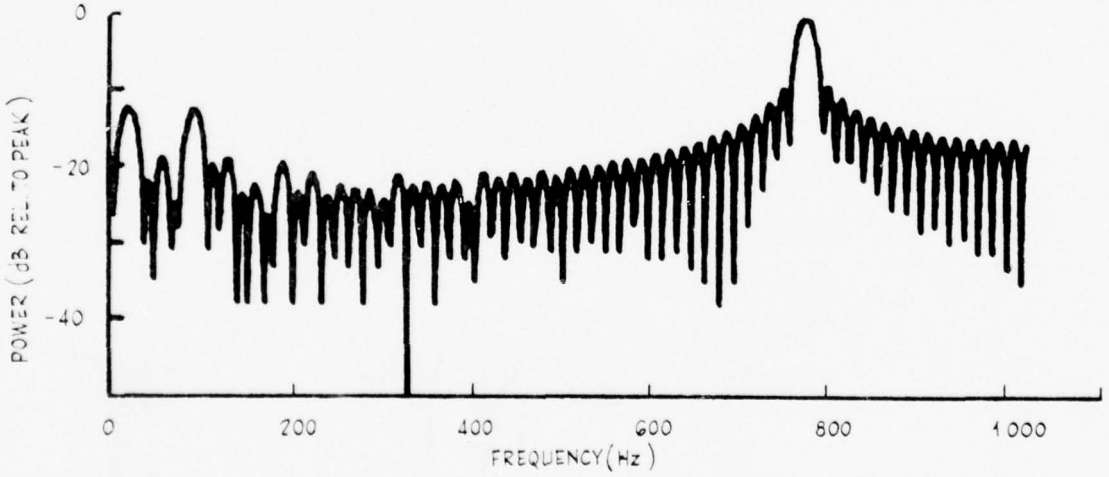
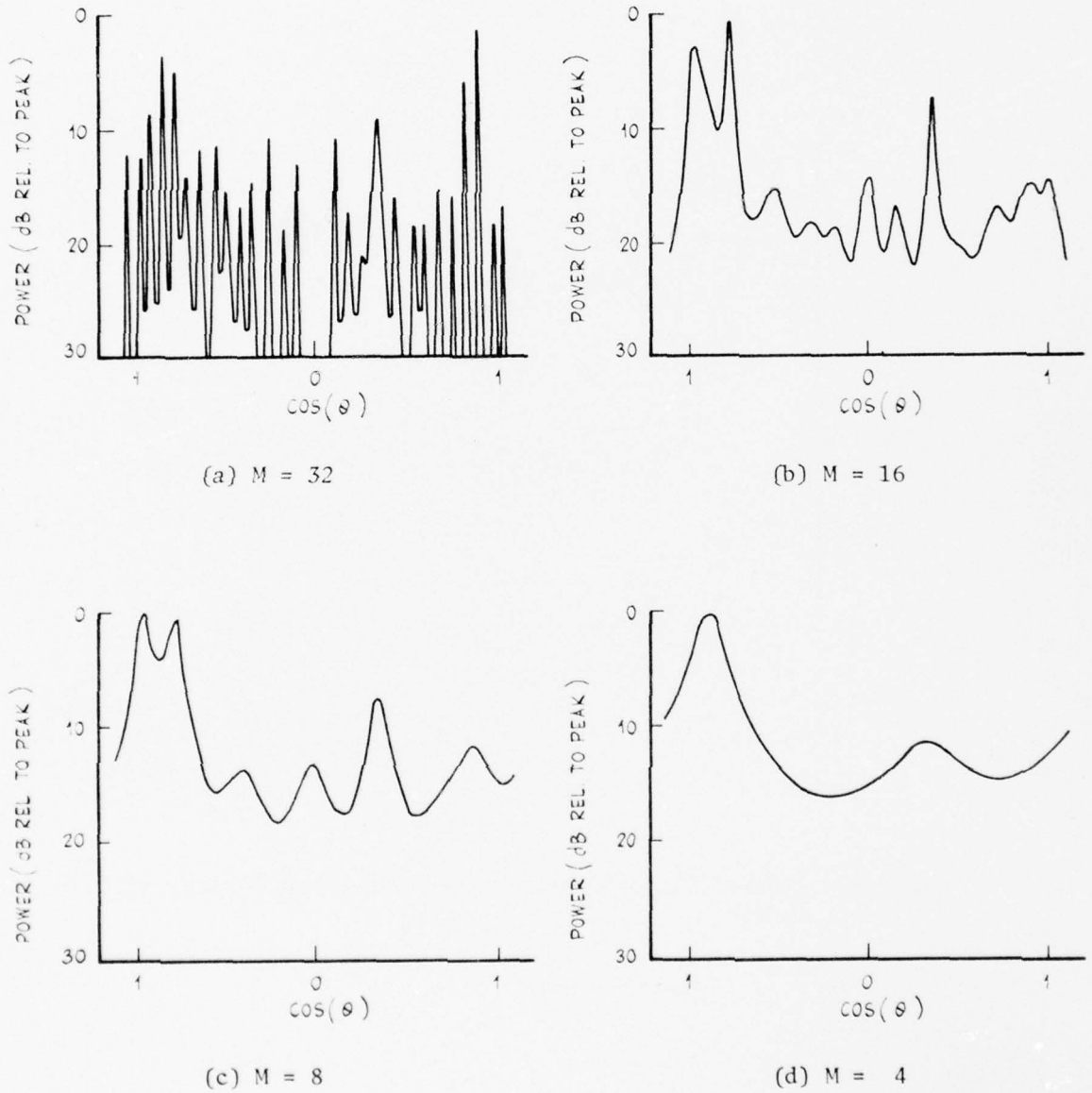


Figure 9(b). Fourier spectrum using 64 lags of figure 8(a)



(a) M = 32

(b) M = 16

(c) M = 8

(d) M = 4

Figure 10. Maximum entropy wave-number spectrum at 0.56 λ using M lags of spatial autocorrelation function

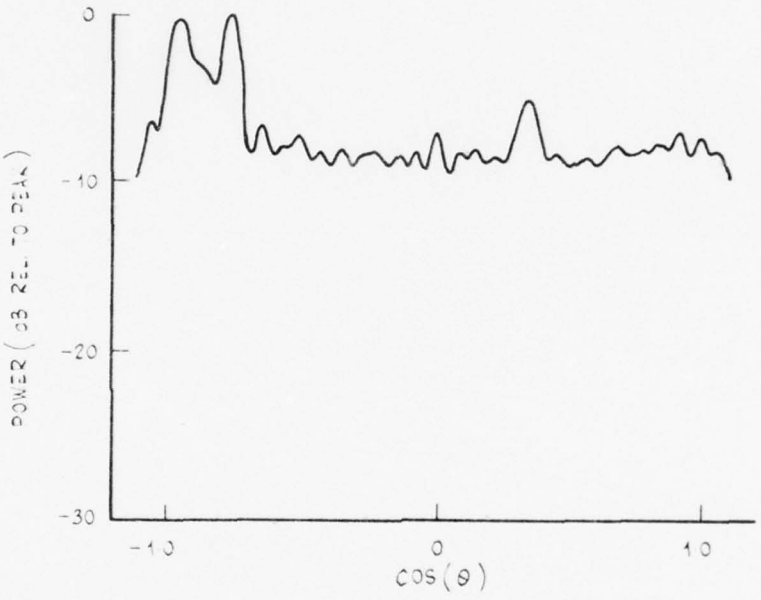


Figure 11. Conventional wave-number spectrum using 32 lags at 0.56 λ

A FLEXIBLE TIME DOMAIN BEAMFORMER

P.C. Drewer, Weapons Research Establishment, Salisbury

SUMMARY

A processor architecture is described which enables beams to be formed with data from a variety of acoustic arrays. While a general purpose computer satisfies the requirement for flexibility, it is shown that the speed limitation is particularly severe. However, by the choice of a suitable architecture based on the operations required, and by performing operations in parallel rather than sequentially, a significant speed increase is possible, while still retaining a large measure of independence from the array geometry.

1. INTRODUCTION

The principle of time-domain beamforming is well known; viz., that if a plane wave front is incident upon an array of receiving elements, then a set of time delays can be calculated for any direction of signal arrival which, when applied to the receiver outputs, causes them to be in phase and to reinforce when added together. The resultant angular response to signals arriving from other than the nominated direction is then a function of the array geometry, relative to the signal wavelength, and any weighting factors which may be applied to the receiver outputs. The effect is to generate a main receiving beam in the desired direction, with a series of undesired subsidiary sidelobes whose magnitude can be controlled to some extent by the choice of a suitable array geometry and the use of some amplitude weightings on the receiver outputs. The direction of arrival of a signal can be determined by simultaneously generating, in a processor, a succession of overlapping main beams, thus giving a measure of the received signal intensity as a function of azimuth.

Variable delays, required for the beamforming process, are difficult to implement using analog techniques. However, if the receiver outputs are sampled and digitized, time delays in increments of the sampling interval are readily obtainable. Other advantages which follow from the use of digital techniques include increased accuracy and freedom from drift. Conceptually, the digital equivalent of a variable delay line would be a shift register (or a series of shift registers, one for each bit in the data sample), with each stage selectable by logic so as to enable the required delay to be implemented. The complexity of this approach is considerable, not the least problem being that of physically accessing each shift register stage.

2. A FLEXIBLE BEAMFORMER

Practical arrays frequently possess some form of symmetry, and for these it may be possible to devise a simplified processor which takes advantage of the symmetrical properties. However, such a processor would not in general be of any use for other arrays. It was therefore felt that there was a need for a processor which could form beams using data from a variety of arrays either already in existence or being contemplated, and which therefore could not depend on using array symmetry. In addition the processor should be able to handle signals with multilevel quantization and to provide array shading. Although the processor was primarily intended for processing hydrophone data, it would be applicable to a wide variety of low-frequency arrays.

More detailed requirements of the processor were as follows:

(a) Data bandwidth to be limited to 2 kHz, with a sampling rate of 8 kHz, giving a maximum phase error (due to time quantization) of $\pm 45^{\circ}$ at the highest frequency.

(b) The number of receivers to be 32 or less.

(c) The signal propagation time across the array to be 32 sample intervals or less.

(d) The number of beams to be formed to be less than 64.

(e) The data precision to be typically 4 bits, but capable of expansion.

Apart from speed requirements, the obvious candidate for a flexible processor is a general purpose computer. By allocating one sample per word, the data precision requirement could be satisfied, even after beamforming, by a 16-bit minicomputer. Rather than attempting to simulate the shift register effect

directly by moving the data within the computer, it would be preferable to set up a "circular" buffer of 32 samples per receiver (i.e., 1K samples total), wherein the most recent receiver sample would overwrite the oldest sample in the appropriate buffer. The delays are then relative to the current value of the buffer pointer (which is common to each receiver), the computation being carried out modulo 32 (the length of each buffer).

In addition to storing 32 samples from each receiver, a table of delays would need to be stored whose size is the product of the number of receiver outputs used to generate each beam and the number of beams formed, i.e., 2K words maximum.

The steps required to generate one beam are as follows:

- Fetch the delay
- Add the pointer
- Mask out all but lowest 5 bits
- Add receiver base address and fetch data
- Add to beam accumulator
- Step on to next base address
- Test for steps completed

The number of machine cycles required depends somewhat on the processor architecture; e.g., the number of working registers, auto-increment and indexing facilities, etc. For a reasonably powerful machine, this could be done in ten cycles. This has to be repeated for each receiver and for each beam formed, i.e., a total of about 20,000 cycles. Although a new frame of 32 receiver samples appears every 125 μs, it is only necessary to form a set of beams every 250 μs. (The 8kHz sampling rate is only required to reduce the time quantization effects, but so far as the information content of the signals is concerned, this is preserved by generating beams at a 4 kHz rate if the receiver outputs have been previously bandlimited to 2 kHz). Thus it can be seen that a cycle time of around 12 ns would be required, which is about 25 to 100 times faster than most currently available small computers.

3. A HARDWARE CONFIGURATION

In a general purpose computer the steps in the beamforming operation referred to above are carried out sequentially. However, in a hardware implementation, they can be performed in parallel. If a "pipeline" structure is used, a full 120 ns is available for each step.

Figure 1 is a simplified diagram of a hardware processor, but included is provision for a shading multiplier for weighting the individual receiver outputs prior to beamforming. Assuming that the data store contains 32 previous samples from each of the receivers to be processed, and that the delay table has been preset with the information required for beamforming, the operation is as follows:

(a) At the end of the previous beam operations, the accumulator is strobed into the output buffer and the accumulator is then cleared.

(b) The step counter is incremented by the system clock and the next control word is read out of the delay table. (As well as providing the delay, the control word supplies the shading multiplier to be applied, the current receiver being accessed, an indicator as to whether this step is the last step in the formation of a beam, and a frame end indicator to show whether this was the last beam to be formed on the current frame of

data. The current receiver information may be required, as in some systems the actual receiving elements used for beamforming may be a function of frequency. If this option is not required, the current receiver address bits can be replaced by a counter.)

(c) The delay is subtracted from the current sample pointer, and the result, modulo 32 (i.e., 5 bits) is combined with the current receiver address (5 bits) to give the 10-bit data address.

(d) The sample selected is read from the data store, multiplied by the shading multiplier, and added to the accumulator, which is building up the beam output.

(e) The steps (b) to (e) are repeated until the "beam-end" bit is set in the control word, whereupon step (a) is repeated, and the beam counter incremented.

(f) Successive beams are then formed by repeating the previous steps, until the "frame end" bit appears in the control word, signifying that all beams have been formed, and that the processor is ready for the next frame of data. The beam counter and step counter are also reset at this stage, and the current sample pointer is incremented.

The loading of the next frame of data can now take place, using much of the logic already provided. The frame loaded actually consists of two sets of receiver samples, the beamforming operation being necessary at only half the data sampling rate. It is assumed that the data frames are externally buffered. The loading operation takes place as follows. Beam counter state zero is recognized as a "new frame" mode and the data store is set into the write mode. The current sample pointer was previously incremented by two relative to the delay selection logic (the current sample pointer being the upper 4 bits of a 5-bit word), and two samples per receiver are written in (with successive delays of 1 and 0). (The current receiver and least significant bit of the delay, i.e., 6 bits total, are available for addressing the external buffer.) After 64 steps of the step counter, the loading is complete, the beam counter is incremented to 1, the data store reverts to the read mode, and beamforming recommences.

The delay table can be similarly set up by forcing it into the write mode, supplying the delays externally, and stepping through with the step counter. In effect the contents of the delay table can be regarded as a primitive form of program supplying instruction words to the rest of the processor to enable it to process signals from the particular array in question.

It can be seen that the use of random access memory in place of shift registers for delays has circumvented the variable tap point problem, as access to all locations is implicit in the nature of the devices. The word length of the data store is determined by the sample precision required, and can be varied readily. Five bits extra precision should be allowed in the adder/accumulator network to allow for growth of the results. It can also be seen that some simplification has been introduced in the addressing logic (mainly the modulo arithmetic) by making the number of samples stored per receiver a power of 2.

For reasons of simplicity, the processor block diagram does not indicate the presence of buffer registers required for "pipeline" processing, but rather it implies that the total of all operations, starting with the incrementing of the step counter and ending with the updated beam total in the accumulator, take place within one clock interval. In practice this is not necessary. By providing buffer registers in appropriate places, each operation, e.g., incrementing, additions, multiplications, memory accesses, etc., can each be

allowed to take up to the full system clock period, the final result appearing several clock intervals later than would otherwise be the case, but still maintaining the same overall processing rate. In effect each step then proceeds in parallel rather than sequentially.

4. IMPLEMENTATION

A prototype processor has been constructed by Mr M. Heigl*. Although there are minor implementation differences, the general philosophy follows that outlined in this paper.

5. CONCLUSIONS

A general purpose computer is unsuitable for beamforming signals from acoustic arrays with many elements in real-time. However it is possible to make a compromise between the flexibility of a general purpose computer and the speed and economy of a hardwired processor designed for a particular array. For relatively little increase in complexity, it is possible to build a processor capable of handling data from a variety of arrays.

* The processor was demonstrated at the Symposium.

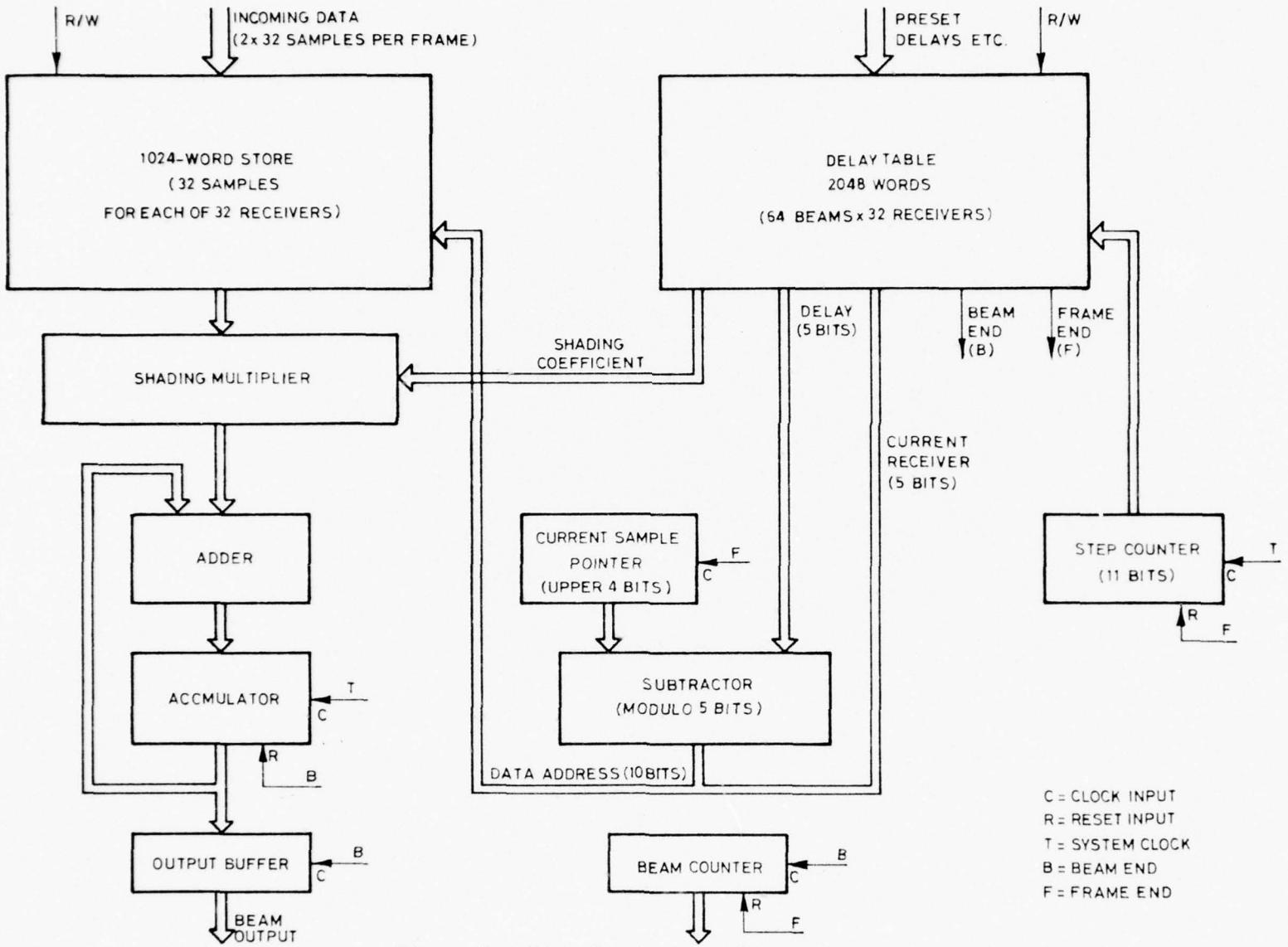


Figure 1. Time domain beam former

WRE-MISC-1
Figure 1

TROPOSPHERIC MULTIPATH PROPAGATION ANALYSIS USING A MICROWAVE HOLOGRAPHIC ARRAY

J.M. Burton and J.V. Murphy, Australian Telecommunications Commission,
Research Laboratories, Melbourne, Victoria

S U M M A R Y

The paper describes an experiment to measure multipath propagation on a line-of-sight microwave radio path. The instrumentation employs a vertical array illuminated by distant and reference transmitters as in holography to determine the amplitude and angle of arrival of the component rays.

The analysis algorithm interprets the received field as the solution to a differential equation to which the data is fitted by minimum mean-square-error approximation. It is then possible to solve the corresponding algebraic equation whose roots are the required angles of arrival of the component rays: the amplitude can be determined by functional analysis. Simulations indicate significantly better resolution than conventional techniques.

1. INTRODUCTION

Microwave radio systems form a substantial part of the Australian Telecommunications Commission trunk network. To date this network has been based on the use of frequency-modulation frequency division multiplex systems.

In the future however, an increasing proportion of overseas networks(ref.1) is expected to be digital. Because of the continuing development of such systems by manufacturers vis-a-vis the more frozen state of analogue system design, such systems may eventually prove cheaper.

When planning microwave systems, propagation data are required to assess the system reliability. The more comprehensive the data, the less over-designed the system needs to be to meet specified performance objectives.

Fading due to rainfall, subrefraction or angle-of-arrival variations increases the noise in an analogue system; this in turn causes an increase in error rate in a digital system. Moreover, line-of-sight microwave links with path lengths greater than 10 km can suffer from multipath fading effects(ref.2) due to the existence of a finite number of simultaneous propagation paths between transmitter and receiver. These extra signal paths distort FM/FDM signals; in the digital case the resultant intersymbol interference increases the error rate - by an amount dependent on the carrier-to-noise ratio.

Some telecommunications administrations propose the transmission of co-frequency, orthogonally-polarised signals using antennas with high cross-polarisation discrimination to double the capacity of digital systems. Variations in angle of arrival and multipath propagation will cause mutual interchannel interference. The statistical occurrence and severity of these effects is therefore of some interest to radio system designers. Previously, measurements from antenna arrays have contributed information about propagation mechanisms(ref.3,4). Vertical radar sounding(ref.5) and spatial correlation studies(ref.6) have also provided answers to fundamental questions such as propagation path microstructures.

The Research Laboratories (Telcom Australia) are implementing a program at X-band to investigate subrefraction and multipath propagation on a line-of-sight path. The program consists of two interleaved experiments:

(a) Data from a vertical array will be analysed by computer to find the amplitudes and angles of arrival of the multipath components.

(b) The X-band signal will be swept over 1 GHz to determine the time delay structure of the multipath components.

The signals from the vertical array will be coherently detected with sufficient phase stability to allow determination of absolute vertical angles and hence calculation of the cumulative distribution of subrefractive fading and angle of arrival variations. The echo delays, which cause distortion or intersymbol interference, will be obtained by processing of the frequency sweep data. In order to remain within the scope of the symposium, only the array experiment will be discussed further, although the signal processing techniques for both experiments are identical at the array.

2. EQUIPMENT

2.1 Design considerations

Similar to holography, the experiment employs a large vertical array which is illuminated by waves from reference and distant transmitters, as illustrated in figure 1. Waves from the distant transmitter can arrive at the array by various paths and an interference pattern or hologram is formed across the receiving aperture. The complex field distribution is then coherently detected and recorded for subsequent computer analysis. As described in Section 3, the analysis algorithm determines the amplitude and angle of arrival of each component wave. The design incorporates the following features:

(i) A reference transmitter is used to reduce phase shift caused by tower motions(ref.3);

(ii) The first mixer associated with each receiving channel obtains its local oscillator drive from the reference transmitter;

(iii) Adaptive scanning of the array using an updated reference element eases oscillator coherence requirements.

The test path length d should be in the range $20 \leq d \leq 50$ km in order to obtain experience relevant to normal system paths. A horizontal path is desirable to allow angles of incidence at elevated layers to remain ≤ 10 mrad. This enhances the possibility of multipath mechanisms, since it increases the reflection coefficient of layers of given strength and ensures that a continuous distribution of delays down to zero delay will be obtained. A constant ground reflection coefficient, as assumed by the computer analysis, is required over the effective reflection zone. These points imply a path over level, uniform terrain or over the sea. Receiver site foreground reflections should be below the system noise level to minimise unwanted interference components across the aperture.

These requirements are basically satisfied by the Pakenham - The Gurdies path. Figure 1 shows the profile of the path and figure 2 shows the location.

A record of the amplitude and phase of the interference pattern along the aperture is obtained from the vertical array. The distance δ between the array elements complies with the sampling theory of bandlimited signals, where to avoid aliasing errors(ref.7),

$$\delta \leq \frac{1}{2 f_{max}} - \frac{\lambda}{2 \sin \phi_{max}} \qquad (1)$$

where f_{max} = maximum spatial frequency (cycles/m),

λ = wavelength (0.027 m)

and ϕ_{max} = maximum angle of arrival measured from the horizontal axis (see figure 1).

Taking ϕ_{max} as 13 mrad, then $\delta \leq 1.03$ m.

For a distant object, the angular resolution is(ref.7):

$$\phi_o \simeq \sin \phi_o = \frac{\lambda}{CL} \qquad (2)$$

where C = resolution improvement factor (see Section 3), assumed $\simeq 10$,
and L = dimension of the receiving array (m).
Choosing ϕ_o = 0.2 mrad we find L = 15.4 m.

A receiving array consisting of 16 sensors with 1 m spacing satisfies equations (1) and (2).

Tower motion produces an erroneous phase shift across the aperture. To overcome this(ref.3), a reference signal is transmitted line-of-sight from a source 1.5 km in front of the receiving aperture. This short path reference signal is quite stable compared to the signal from the distant transmitter. It can be shown that the phase error due to the motion of a receiving sensor is:

$$\Delta\phi = \Delta d(\underset{\sim}{k}_r \cdot \hat{n} - \underset{\sim}{k}_s \cdot \hat{n}) \qquad (3)$$

where Δd = horizontal change in sensor position,

$\underset{\sim}{k}_r$ = reference wave vector,

$\underset{\sim}{k}_s$ = signal wave vector,

and $\hat{n}$ = unit vector normal to the array.

With the reference transmitter set up as shown in figure 1, a mast movement of Δd = 0.1 m at the top causes a phase error of 44 mrad, which is considered acceptable.

If two high quality 5 MHz quartz oscillators are used as sources, electronic scanning eliminates the need to phase lock the Pakenham transmitter to the reference. In this case, the long term drifts are relatively unimportant, and the short term fluctuations are the limiting factor. From Van Trees (ref.8), the elapsed phase difference between the two signals at X-band, $\Delta\theta$, can be modelled as a simple Wiener process, i.e.,

$$\text{var.}(\Delta\theta) = t/\tau \qquad (4)$$

where var. denotes variance, t = scanning period, and τ = coherence time. For $\Delta\theta \leqslant 90$ mrad and τ = 0.5 s, it follows t $\leqslant$ 4 ms.

2.2 Implementation

Figure 3 shows the block schematic of the reference transmitter. The configuration of the Pakenham transmitter is similar except for its lower ERP (66 dBm) and 10 kHz frequency offset. Measurements of short-term frequency deviation at 10.8 GHz and residual FM noise at a 10 kHz offset from the carrier, have shown that negligible additional frequency fluctuations are introduced by the multiplication chain.

Initially the vertical receiving array will consist of 16 X-band horns mounted on the top 16.5 m of a 76 m mast. Each horn has a gain of 23 dBi, and the half power beam widths are about 13^o. Vertical polarisation is used to reduce unwanted foreground reflections near the Brewster angle (ref.9,10). Horns have the advantages of being relatively cheap and simple to manufacture.

Figure 4 illustrates an array element receiving channel. Each sensor horn feeds a two-port waveguide "starved mixer". The reference signal acts as a low level local oscillator drive offset 10 kHz from the Pakenham transmitter. DC bias is used to improve the mixer's performance at the expense of noise figure. The amplified IF signals are then coherently detected. The high-quality 10 kHz oscillator is tuneable allowing long-term frequency changes in the 5MHz standards to be 'nulled' out. Scanning of the

array is under the control of a minicomputer which selects the required in-phase and quadrature channels to be digitised and recorded. To reduce phase errors across the aperture over a scan period, the minicomputer selects the sensor with the highest output level from the previous scan. During the current scan, the scan order is then:

$$S_1, S_r, S_2, S_r, \ldots, S_{16}, S_r,$$

where S_i denotes the ith sensor, and S_r denotes the reference sensor chosen. Differential phase values are then calculated relative to the reference sensor. Figure 5 illustrates the 10 kHz phase error due to mast movement, oscillator phase fluctuations and coherent detector errors.

3. DATA PROCESSING

Together with housekeeping information, the data is recorded on magnetic tape for subsequent processing. Let the complex voltage measured at height x on the array be denoted by $v_m(x)$, where the time-dependence is suppressed. Then if multipath propagation occurs $v_m(x)$ must be of the form

$$v_m(x) = \sum_{k=1}^{K} r_k \, e^{jx\xi_k} + w(x) \qquad (5)$$

where $\xi_k = \frac{2\pi}{\lambda} \sin \phi_k$,

x = distance from bottom of array,

λ = wavelength,

r_k = complex amplitude of kth wave,

ϕ_k = angle of arrival of kth wave

and w is a random process. In fact the sample functions from the random process will have as their domain the measurement period of the experiment, i.e., $x = x(t)$.

The measured array distribution can firstly be filtered to reduce the effect of added noise, thus,

$$v(x) = \int_0^L h(x,y) \, v_m(y) \, dy \qquad (6)$$

where L = array length. The optimum filter characteristic h can be determined (ref.8) provided the power spectrum of the signal is known. In practice this involves assigning relative probabilities of occurrence to the various angles of arrival. The object of multipath analysis is to estimate K and the set $\{r_k, \xi_k\}$ by M and $\{a_m, \theta_m\}$; this involves the approximation of v by the finite linear

combinations of functions $e^{j(.)\theta_m}$:

$$v = \sum_{m=1}^{M} a_m e^{jx\theta_m} . \tag{7}$$

Using the method of Diamessis(ref.11), equation (7) is recognised as the solution of the differential equation

$$\frac{d^M v}{d(jx)^M} + b_{M-1} \frac{d^{M-1} v}{d(jx)^{M-1}} + \ldots + b_o v = 0 \tag{8}$$

and the θ_m are found as the solutions of the corresponding algebraic equation:

$$\theta^M + b_{M-1} \theta^{M-1} + \ldots b_m \theta^m \ldots + b_o = 0 . \tag{9}$$

By integrating equation (8) M times the approximation is transformed into the form:

$$v \approx \sum_{m=0}^{M-1} d_m P_m - \sum_{m=0}^{M-1} b_m v^{(M-m)} \tag{10}$$

where

$$v^{(m)}_{(x)} = \int_0^{jx} \int_0^{jx_m} \ldots \int_0^{jx_2} v(x_1)\, d(jx_2) \ldots d(jx_m)$$

and P_m is the monomial

$$P_m = (jx)^m/m!.$$

Since θ_m is defined to be real, b_m must also be real. Then the functions defined by the RHS of equation (10) form a convex set κ in the Hilbert(ref.12) space $\mathcal{L}_2(\mathcal{J})$, $\mathcal{J} = [-L/2, L/2]$. Let k_o be the vector in κ nearest to v. Then by an analogous argument to Luenberger(ref.13) it can be shown that in a complex Hilbert space the optimum approximation v_o satisfies

$$\mathrm{Re}(v - v_o, v' - v_o) \leq 0, \quad v' \in \kappa,$$

where the inner product in $\mathcal{L}_2(\mathcal{J})$ is defined by

$$(u,v) = \int_{-L/2}^{L/2} u(x')\, v^*(x')\, dx', \quad u,v \in \mathcal{L}_2(\mathcal{G}),$$

and x' is the distance from the array centre.
Let

$$v' = v_o - v^{(n)};$$

then

$$\mathrm{Re}(v - v_o, v^{(n)}) \geqslant 0;$$

thus from equation (10)

$$\sum_{m=0}^{M-1} \mathrm{Re}(d_m P_m, v^{(n)}) - \sum_{m=0}^{M-1} b_m \mathrm{Re}(v^{(M-m)}, v^{(n)}) = \mathrm{Re}(v, v^{(n)}),$$

$$n = 1, \ldots, M. \qquad (11(a))$$

From the projection theorem(ref.13) we have for the P_m,

$$\sum_{m=0}^{M-1} (d_m P_m, P_n) - \sum_{m=0}^{M-1} b_m (v^{(M-m)}, P_n) = (v, P_n),$$

$$n = 1, \ldots, M. \qquad (11(b))$$

Equations (11) are equivalent to a 3M x 3M real matrix equation compared with the 4M x 4M dimension of the normal equations. When the $\{b_m\}$ have been found the set $\{\theta_m\}$ will be found(ref.11) as the solution of equation (9). Then the matrix $\underline{A}$ of coefficients $a_1 \ldots a_M$ which result in the minimum mean-square-error approximation is given by(ref.13,14) the solution of:

$$\underline{\Lambda}\,\underline{A} = \underline{S} \qquad (12)$$

where $\underline{\Lambda}$ and $\underline{S}$ have elements

$$\lambda_{m,n} = (e^{j(.)\theta_n}, e^{j(.)\theta_m})$$

$$= \frac{\sin L(\theta_m - \theta_n)/2}{(\theta_m - \theta_n)/2} \qquad (13)$$

and

$$a_m = (v, e^{j(.)\theta_m}) = \int_{-L/2}^{L/2} v(x)\, e^{-jx\theta_m}\, dx.$$

The amplitudes a_m must then be adjusted according to the inverse of the filter characteristic in equation (6).

3.1 Dependence of approximation error on number of components

When the number of components is unknown a reasonable strategy is to start at M = 1, 2 ... and monitor the error at each stage. Let $\{\theta_m, m=1, M\}$ be any set of angles; then the corresponding approximation error is

$$\epsilon(\theta_1, \ldots, \theta_M) = v - \sum_{m=1}^{M} a_m e^{j(.)\theta_m}$$

where the a_m are found from equation (12). The norm of the error is(ref.14)

$$\|\epsilon\|^2 = \|v\|^2 - \underline{A}^{T*}\, \underline{S}. \qquad (14)$$

Clearly $\min_{\theta_1,\ldots\theta_m} \|\epsilon\|$ is a non-increasing function of M. Since there is a significant amount of effort in solving equations (9), (11) and (12) it may be worthwhile before progressing from the Mth to the (M+1)th stage, to obtain an upper bound for the consequent improvement in the error. This bound will also indicate the onset of error build-up which will occur if $M > K$ or $\xi_k \approx \xi_\ell$ for some k, ℓ.

The bound can be derived by considering a suboptimal procedure for progressing from the Mth to the (M+1)th stage. For any v in $\mathcal{L}_2$, given $\{\theta_m, m=1,\ldots,M\}$ as the roots of equation (9), the (M+1)th component can be chosen as the projection of $\epsilon(\theta_1,\ldots,\theta_M)$ along $e^{j(.)\theta_{M+1}}$. Then

$$a_{M+1} = (\epsilon(\theta_1, \ldots, \theta_M), e^{j(.)\theta_{M+1}})/L$$

and(ref.13)

$$\|\epsilon(\theta_1, \ldots, \theta_{M+1})\|^2 = (\epsilon(\theta_1, \ldots, \theta_M) - a_{M+1} e^{j(.)\theta_{M+1}}, \epsilon(\theta_1, \ldots, \theta_M)).$$

Hence the optimum error at the (M+1)th stage must satisfy

$$\| \min_{\theta_1,\ldots\theta_M} \epsilon(M+1)\|^2 < \|\epsilon(\theta_1, \ldots, \theta_M)\|^2 - \max_{\theta} |(\epsilon(\theta_1, \ldots, \theta_M), e^{j(.)\theta})|^2/L$$

$$= \|v\|^2 - \mathbb{A}^{T*}\mathbb{S} - \max_{\theta} \left| (v, e^{j(.)\theta}) - \sum_{M=1}^{M} a_m \frac{\sin L(\theta - \theta_m)/2}{(\theta - \theta_m)/2} \right|^2, \qquad (15)$$

where ^{T*} stands for matrix transpose, complex conjugate respectively. At each stage the vector product in equation (15) is incremented by adding one term and a unidimensional search, the precision of which contributes to the tightness of the bound, is performed.

3.2 Simulation

To test the feasibility of the data processing technique a simulated signal of the form equation (5) was generated by computer. The filter characteristic needs to be limited in the x-domain to avoid excessive edge-effect errors in evaluating the convolution equation (6). Accordingly the characteristic proposed by Papoulis(ref.15) which is optimum in another context was chosen. This window function obeys

$$h(x) = 0, \quad |x| > X$$

and the half-power width is also approximately X. The error in determining the component amplitudes and angles of arrival is given by

$$\epsilon = \sqrt{\sum_{m=1}^{M} \left(\frac{|a_m| - |r_m|}{|a_m|} \right)^2 + \sum_{m=2}^{M} \left(\frac{[\theta_m - \theta_1] - [\xi_m - \xi_1]}{\xi_m - \xi_1} \right)^2}$$

when M is set equal to K. As X/L increases the noise power is reduced and ϵ consequently decreases until effective shortening of the record due to the finite width of h in the convolution (6) becomes apparent. This behaviour is shown in figure 6 for the case of a 64-element array. The minimum of the curve occurs for smaller values of X/L than would be expected from the spectra of the signal and noise. Further refinement of the numerical integration procedures may overcome this discrepancy. The dependence of the error on signal-to-noise ratio is shown in figure 7. It can be seen that the angle of arrival of the second component is well within the array beamwidth so that the resolution is really related to the acceptable error in the parameter determination. The resolution clearly depends on the signal to noise ratio but values of 12 for 2 and 6 for 3 components have been demonstrated.

4. DISCUSSION

The paper has described an experiment currently being implemented using a linear coherent array operating at one frequency and referred to a complementary experiment employing a single receiving antenna and a wideband test signal. The objective is to obtain a better model of the troposphere for radio system planning purposes. The specific data required for an adequate model is the

statistical distribution of multipath wave amplitudes and arrival angles. The core of the experiment is the processing of the complex received signal vector. Using conventional Fourier analysis the resolution in angle of arrival (or delay) is of the order of the inverse array length in wavelengths (or signal bandwidth). The method of data processing described improves this figure by about an order of magnitude and it is hoped that the resolution will be sufficient to determine, by comparing arrival angles and delay times, whether multipath component waves are continuously refracted by thick tropospheric layers of high refractive index gradient or sharply reflected by thin layers.

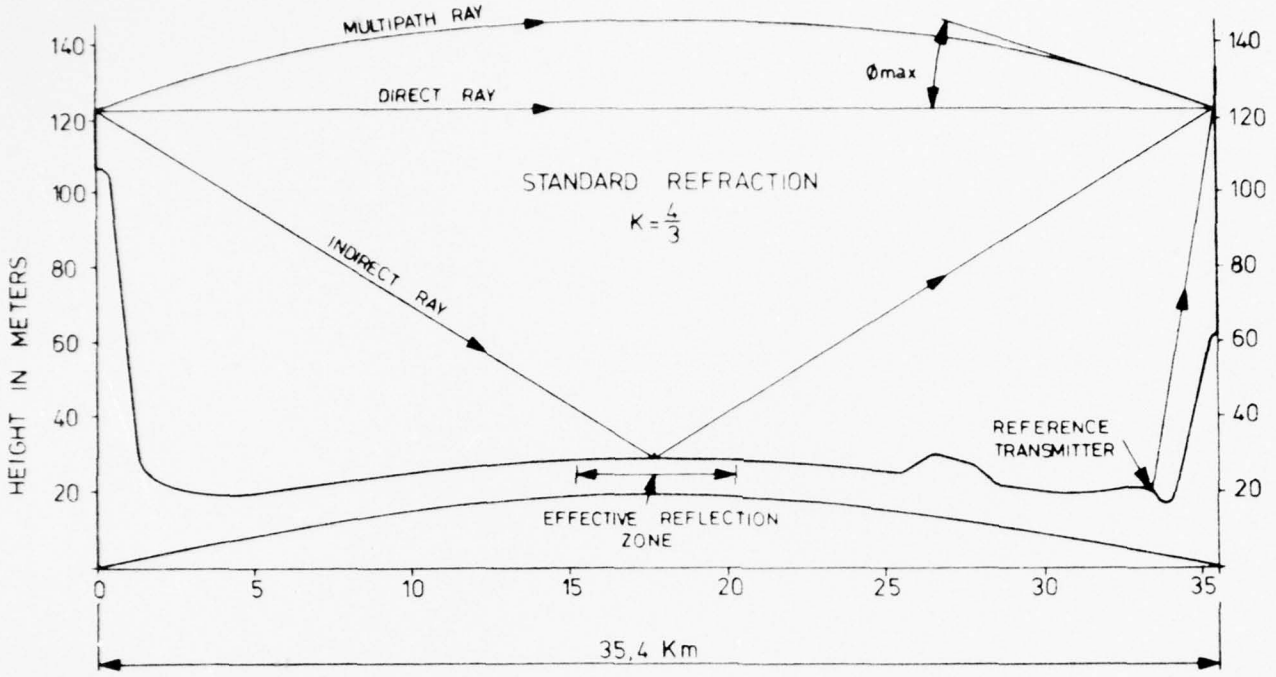
5. ACKNOWLEDGEMENT

We wish to thank Mr. B. Milne of Monash University for developing and programming the numerical procedures. The permission of the Australian Telecommunications Commission to present this paper is gratefully acknowledged.

REFERENCES

No.	Author	Title
1	Harden, B.N. and Turner, D.	"Propagation Studies and the Development of Terrestrial Microwave Radio Relay Systems above 10 GHz in the U.K.". IEE Conf. on Prop. of Radio Waves at Frequencies above 10 GHz, London, 1973, pp 1-5
2	Misme, P. Battesti, J. and Boithias, L.	"The respective influences of multi-path configurations and precipitation rates for frequencies lying between 10 GHz and 30 GHz". IEE Conf. on Prop. of Radio Waves at Frequencies above 10 GHz, London, 1973, pp 56-64
3	Cox, D.C. and Waterman, A.T.	"A Data-gathering Antenna Array for Trans-horizon Propagation Research". IEEE Trans. Ant. and Prop., Vol. AP-17, January 1969, pp 69-77
4	Waterman, A.T.	"A rapid beam-swinging experiment in transhorizon propagation". IRE Trans. Ant. and Prop., Vol. AP-6, October 1958, pp 338-340
5	Richter, J.H.	"High resolution tropospheric radar sounding". Radio Science, Vol.4, No.12, December 1969, pp 1261-1268
6	d'Auria, G. and Solimini, D.	"Coherence properties of a 9 GHz propagation path near the ground". Radio Science, Vol.5, No.12, December 1970, pp 1387-1395
7	Metherell, A.F., (Ed.)	"Acoustical Holography". Plenum Press, New York, 1969
8	Van Trees, H.L.	"Detection, Estimation and Modulation Theory". John Wiley and Sons, 1971
9	Barton, D.K.	"Low Angle Radar Tracking". Proc. of IEEE, Vol.62, No.6, June 1974, pp 687-704
10	Attwood, S.S., (Ed.)	"Radio Wave Propagation". Academic Press Inc., New York, 1949, pp 259-268
11	Diamessis, J.E.	"Least-square-exponential approximation". Electron. Lett. Vol.8, pp 454-455, 7 September 1972
12	Simmons, G.F.	"Introduction to topology and modern analysis". New York: McGraw-Hill, 1963
13	Luenberger, D.G.,	"Optimisation by vector space methods". New York: Wiley, 1969

No.	Author	Title
14	Brown Jr., J.L.	"Anharmonic approximation and bandlimited signals". Info. and Control, Vol.10, pp 409-418, April 1967
15	Papoulis, A.	"Minimum-bias windows for high resolution spectral estimates". IEEE Trans. Inform. Theory, Vol.19.1, January 1973, pp 9-12



PAKENHAM	
GROUND LEVEL	: 107m ASL
MAST HEIGHT	: 15 m
ANTENNA	: 1,8m PARABOLA
POLARIZATION	: VERTICAL
FREQUENCY	: 10.8 GHz

REFERENCE	
GROUND LEVEL	: 18m ASL
MAST HEIGHT	: 3 m
ANTENNA	: 1,8m PARABOLA
POLARIZATION	: VERTICAL
FREQUENCY	: 10.80001 GHz

THE GURDIES	
GROUND LEVEL	: 61 m ASL
MAST HEIGHT	: 76 m
ANTENNAE	: 16 x X-BAND HORNS
POLARIZATION	: VERTICAL

Figure 1. Pakenham - The Gurdies path profile

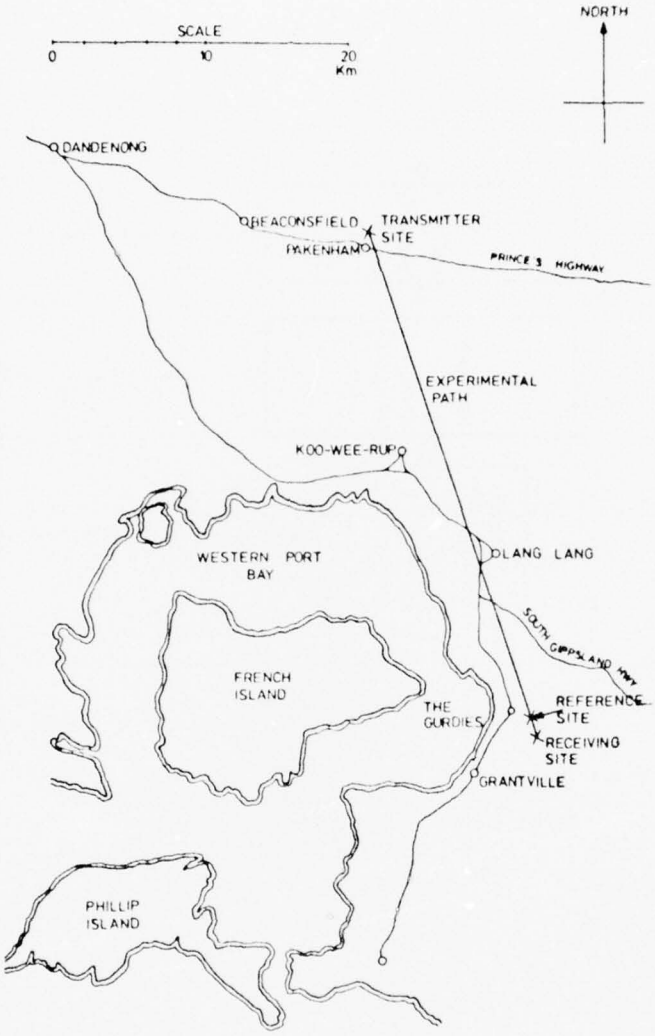
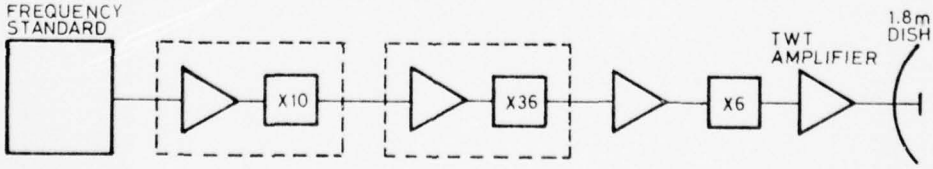


Figure 2. Location of experimental radio path
Pakenham - The Gurdies



POWER dBm	+10	+10	+30	+37	+26	+44
FREQUENCY	5MHz*	0.05 GHz*	1.8GHz*	1.8GHz*		10.8GHz ±10kHz

+85

*FREQUENCY STANDARD IS ADJUSTED TO GIVE 10kHz OFFSET RELATIVE TO PAKENHAM TRANSMITTER

Figure 3. Reference transmitter

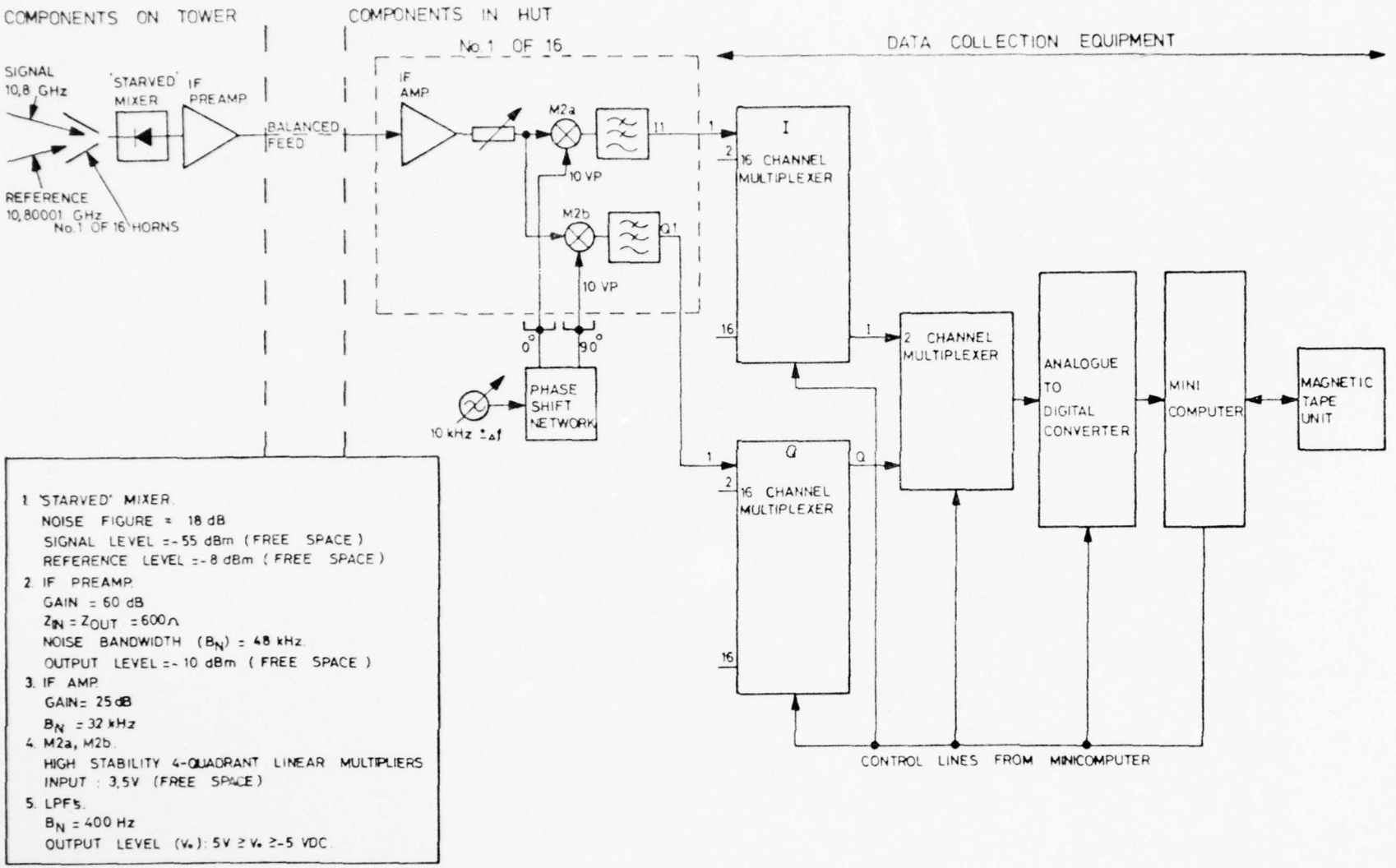


Figure 4. Block schematic of one receiving channel, recording and control equipment

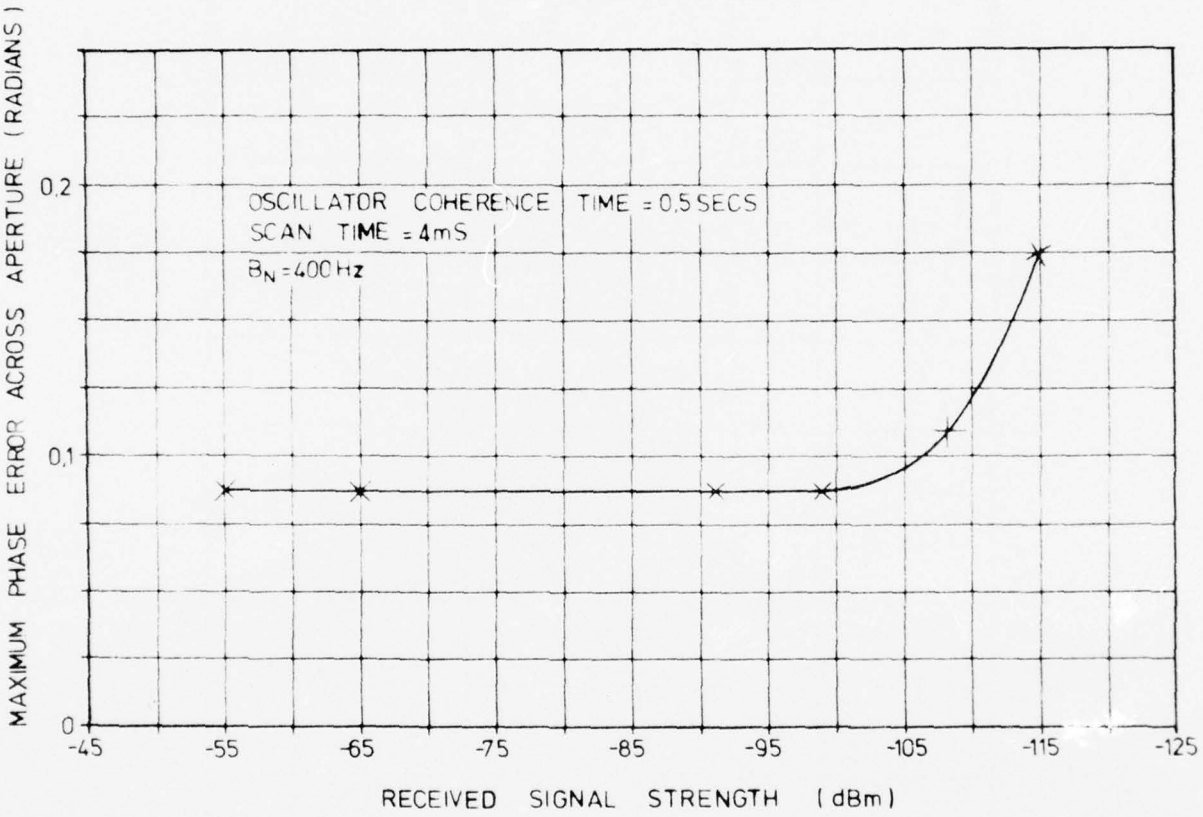


Figure 5. Phase error vs. signal level

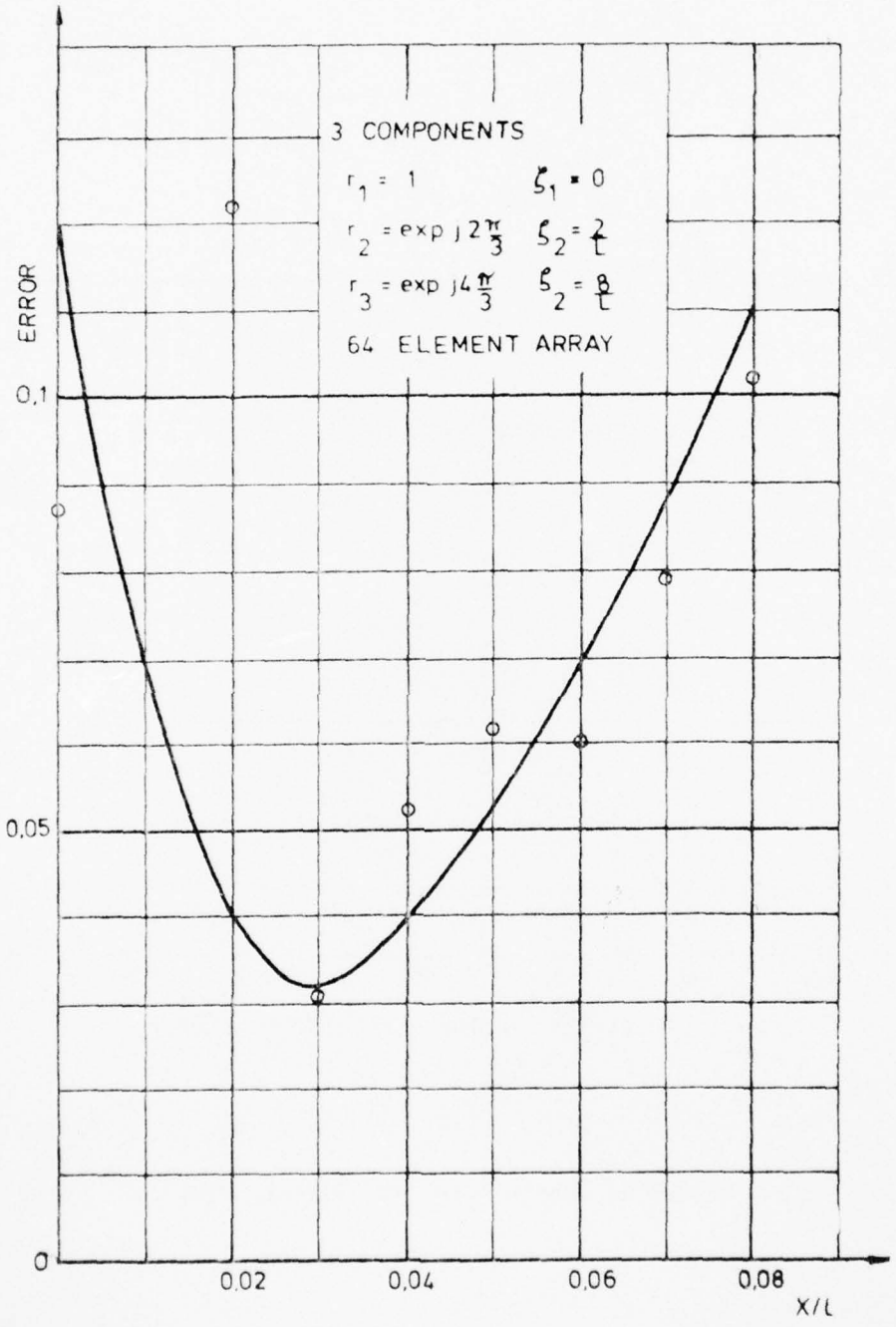


Figure 6. Dependence of error on filter width, SNR = 50 dB

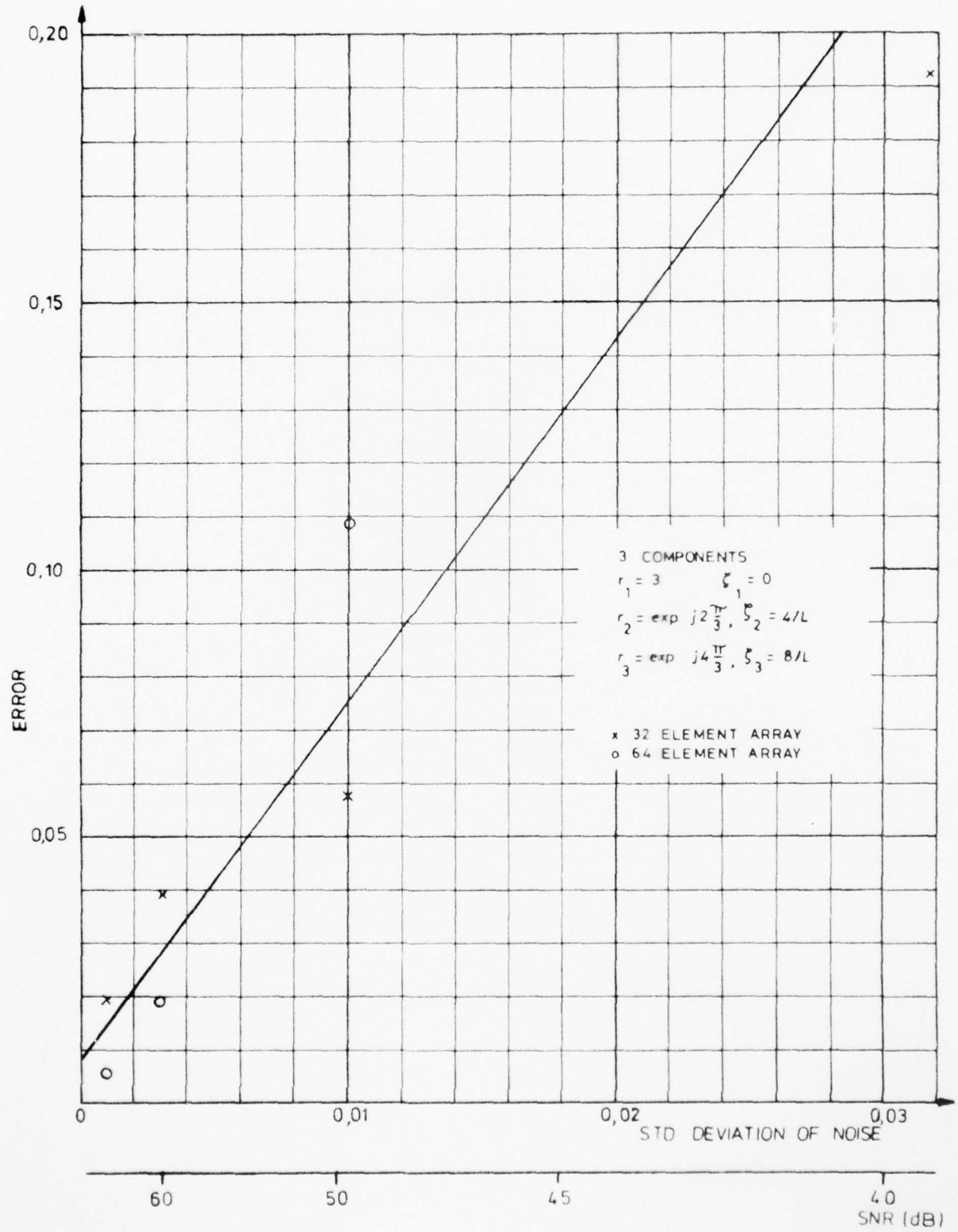


Figure 7. Error in parameter estimation vs. input noise level

SCAR - THE MPL SCATTERING ARRAY

T.G. Birdsall and G.T. Kaye, University of California, San Diego
Marine Physical Laboratory of the Scripps Institution of
Oceanography San Diego, California 92132

S U M M A R Y

A horizontal two-dimensional receiving array of 128 hydrophones arranged in a 45 ft diameter circle about a single source (transmitter) is used to resolve scattering or reflecting planes or points directly below the array in 1024 1-ft depth increments. The purpose is the study of scattering from water density microstructure, such as internal waves.

The signal processing is designed for real-time on-site use with a repetition interval of 5 s. The processing uses hard clipping (1 bit) at the receivers, and a sampling rate of up to 4 kHz. The digital processing successively focuses the array at each of 200 plane zones (each about 1 ft square) at each depth.

The processing hardware is based on a PDP11 minicomputer with disc, and a Microdata 3200 microprogrammed processor.

The design philosophy, implementation, key problems, and operation will be discussed.

1. INTRODUCTION

The Scattering Array (SCAR) is a large-aperture planar hydrophone array with a sound source that is being used to observe the sound scattering from fine-scale, layered structures of water temperature and salinity gradients. From these observations the evolution and destruction of density features in the upper 400 m of the oceanic water column will be studied.

The array and source are mounted on a 70 ton submersible barge. This barge has physical dimensions of approximately 15 m x 15 m x 2 m. It is towed to sea, submerged and suspended from a surface support vessel. Then the sound source and 128 hydrophones are connected to the signal processing hardware aboard the surface support vessel for real-time data monitoring and analysis. When submerged, the barge is a very stable platform which varies less than $\frac{1}{2}^{o}$ from a horizontal orientation.

The basic construction of the array is shown in figure 1. Seven modules of angled steel were constructed and bolted together to form the barge. Within each module are mounted two pipes, 60 cm in diameter and 14 m in length, which serve as buoyancy tanks. Piping runs from the compartments of these fourteen tanks to the diving control station on the bow. The fully assembled array is shown in figure 2. It is loaded aboard a large trailer for towing ashore and is launched via a seaplane ramp. The trailer is detached from the barge after launching. The tubes on top of the barge house the suspension system. On the bottom of the barge (figure 3) are mounted 128 hydrophones with their accompanying cones and baffles. All of the cones lie within the same horizontal plane to $\pm$ 1/3 cm, when the array is in the water. The sound source is shown in the upper centre of the figure.

After the barge has been towed to sea (figure 4), four suspension cables are connected to the surface support vessel. Two SCUBA divers on the barge flood the buoyancy tanks and dive the array to a water depth of 20 m. Then the electrical cable bundle is lifted to the surface and connected to the computer hardware aboard the surface support vessel. The suspension cables are connected to a series of compression springs in the suspension tubes. The movement of the surface vessel is compensated by spring compression rather than by array motion, for vertical excursions of 3 m or less. As a result, the array is effectively decoupled from the motion of the sea surface, thus avoiding a difficult signal processing problem.

A few basic words are in order for the reader who is unfamiliar with underwater acoustics. The speed of sound in seawater varies from about 1500 m/s at the surface to perhaps 95% of this in the cold water of a deep ocean. In the large oceans the speed of sound will increase again at great depths due to the increased pressure, even increasing beyond that at the surface. Such a picture has to be modified to admit a complicated regime of temperature and salinity non-uniformities that produce a layered structure with sharp gradients, moving perhaps randomly or perhaps propagating as gravity driven waves.

The variability of the sound speed is the major source of confusion in all long range acoustic propagation, and even in many short range, high precision situations. On the credit side, the slow speed of sound (think of it as meaning 1,230,000 μs/mile) allows extensive computer processing to occur during the time it takes for the pulse to propagate and return.

The slow speed of sound has another impact on active sonars. Sequential scanning of angles, in two dimensions, would take "forever" if each "look" consisted of one or more pulses transmitted in and received in a narrow beam. Sonars generally transmit in wide beams, but form a large number of "preformed receiving beams" which process the same return. Thus angular resolution is the business of the receiving array, not the transmitter. These receiving arrays usually consist of a large number of hydrophones (point sensors) which may be configured in a line, about a surface, or throughout a volume.

SCAR is a research array, and as such the array and the processing must be robust enough to succeed in spite of many unknowns, and adaptable enough to be modified easily as more becomes known. The adaptation capability means that the signal processing is not cast in concrete. We can report its current state, but it will change.

The basic parameters that the SCAR system should extract from layers or layer boundaries are

(1) where are they?

(2) are they flat or tilted?

(3) is the return from a layer or from point scatterers in the layer?

The first question of "where?" is answered by a depth, meaning straight down. One of the original unknowns was the fluctuation from horizontal that SCAR itself would experience. This was one reason that a large number of beams were selected (16 x 16 = 256 beams) and why auxiliary level sensors were mounted on SCAR. Recent tests have indicated the maximum array tilt is about 0.5^{o}.

Layer tilt will be determined by a combination of measurements from beams at different angles, and from the layer depth plots that will develop as the layers propagate or drift past on currents.

The ocean is filled with many point reflectors or scatterers, biological life with gas or oil-filled bladders (for density control) and various types of debris. A density gradient may cause an echo directly, or perhaps an even larger echo may occur because of life or debris that has temporarily collected at a specific density.

A major research uncertainty is the best operating frequency, or suite of frequencies. The transmitter and processing currently operate from 2 kHz to 16 kHz, using a 0.5 ms to 1 ms pulse. The array beam pattern changes considerably over this 8:1 frequency band. To eliminate secondary lobes each hydrophone is mounted in a cone coated with pressure release material to restrict its effective field of view to $\pm 5^{o}$.

A final and quite common uncertainty is the required dynamic range, since scattering strength from the layers is not well known. The SCAR processing currently uses a polarity-only processing ("hard clipping") which destroys the absolute level but which acts to preserve relative levels "locally". A high degree of redundancy in computation helps to smooth through roughness caused by clipping. That is, the beam and range steps are considerably finer than resolution would dictate.

2. SIGNAL PROCESSING

All 128 hydrophones and auxiliary array housekeeping signals are individually cabled up to the Oceanographic Research Buoy, ORB, from which SCAR is suspended. After amplification each hydrophone signal is quadrature demodulated, which converts the physical real signal to a complex valued analysis signal. This is currently accomplished by taking a pair of samples at one quarter cycle transmit frequency spacing. The timing between pairs of samples is always an integer number of cycles of the transmit frequency and at least 250 μs.